Using GIS to Identify Potential Sample Sites for Assessing the Effects of PL-566 Structures on the Ecological Integrity of Missouri's Streams

A Final Report to the Missouri Department of Natural Resources
under agreement MDNR G02-566-1

By Scott P. Sowa and Robbyn Abbitt

Missouri Resource Assessment Partnership
University of Missouri
4200 New Haven Road
Columbia, MO 65201

March 2004

EXECUTIVE SUMMARY

The Watershed Protection and Flood Prevention Act (PL-566) authorizes the USDA Natural Resources Conservation Service to help local organizations and units of government plan and implement watershed projects. PL-566 watershed projects are locally led to solve natural and human resource problems in watersheds up to 250,000 acres in size. These projects can include flood prevention and damage reduction, development of rural water supply sources, erosion and sediment control, fish and wildlife habitat enhancement, wetland creation and restoration, and increased recreational opportunities. Flood prevention measures include land treatment, and structural or non-structural measures that reduce or prevent floodwater damage by reducing runoff and erosion, or reducing the frequency or severity of flooding. Structural flood-control measures typically involve the construction of numerous headwater impoundments within the project watersheds.

Existing scientific literature suggests that headwater impoundments can have both positive and negative effects on biological integrity. In Missouri, there has been no explicit examination of the potential benefits or impacts of these small flood control structures on biological integrity. As a result of questions raised during the 401/404-permit process an interagency advisory committee was formed to help devise a research strategy to answer many complex questions. The advisory committee agreed that two preliminary objectives must be addressed prior to devising specific research projects to assess any potential effects of PL-566 structures on biological integrity. The first objective entailed a comprehensive review of the existing literature in order to assess existing study designs and research findings of projects that have examined the potential effects of impoundments on the biological integrity of streams. This objective was addressed in a companion project (Doisy and Rabeni 2004).

In order to specifically study the potential effects of PL-566 structures with treatment vs. control or correlative study designs, measures must be taken to ensure that comparisons are only made among sites expected to have relatively similar ecological character in the absence of these structures. To address this issue the advisory committee devised a second objective which involved developing a GIS database that could be used as an initial coarse-filter selection tool to help select study sites with similar watershed landscape and land use character and also assess opportunities or limitations for treatment vs. control and correlative experimental designs. This objective is the focus of this report.

Most PL-566 structures have been constructed in northern Missouri and this was the focus of our project. A 300-cell digital stream network, consisting of 217,860 arcs (i.e., stream segments) was generated for the study area. Various measures of stream size and stream gradient were generated for each segment. Waterbodies (2-15 acres and 2-50 acres) were extracted from the MoRAP land use/cover database and intersected with headwater streams within the digital network. The presence and number of headwater impoundments within the watershed of each stream segment was then calculated. We were unable to specifically identify PL-566 structures, so studies specifically designed to assess these structures will have to rely on additional information sources to isolate these structures from other headwater impoundments.

Watershed percentages for several geology, soil, relief, and land cover variables were then generated for each segment. These data were used to classify segments into relatively distinct groupings (stream types) using multivariate cluster analysis. Cluster analyses were performed on two sets of watershed landscape variables, a) a Full Set that contained 19 geology, soil, and

relief variables, and b) a Reduced Set that contained 8 soil and relief variables. Separate cluster analyses were also performed on two sets of Land Cover variables, a) a set that pertained to the overall watershed of each segment, and b) a set that pertained to the local (immediate) drainage of each segment.

Diagnostic plots were used to assess how many distinct clusters existed within each of the four datasets. Approximately 14 to 18 distinct stream types were identified with the Full Set of variables, while 12 to 16 were identified with the Reduced Set. At least 10 clusters should be used with either of the Land Cover sets. Ten, spatially-related, GIS databases were developed for this project that can be used as an efficient tool for assessing study design options and also identifying and mapping a pool of potential replicate study sites that are relatively similar with regard to watershed landscape character and both watershed and local land cover. Detailed descriptions of these databases are provided as well as instructions on how to collectively use the databases for devising research projects.

Because relatively coarse-scale geospatial datasets were used to generate the cluster groupings for this project and these datasets are certainly not without error, the resulting GIS databases should be regarded as an efficient initial coarse-screening tool for assessing potential study designs and selecting potential study sites. Whenever possible, potential replicates should be selected so that they are geographically situated as close together as possible. Also, additional higher resolution datasets should be used along with field visits to further assess the relative similarity of the initial pool of potential replicate sites.

Finally, we conducted an assessment of opportunities or limitations for treatment vs. control and correlative designs. We were interested in how study design options might change; a) as stream size increased (headwater, creek, small river), b) with the different cluster results (Full vs. Reduced Sets), and c) as the number of clusters within each set increased. For every scenario we examined there was sufficient replication potential for treatment vs. control experiments for headwater streams. As stream size increases, however, the potential for devising treatment vs. control studies diminishes quickly. In no instance, was there an opportunity for devising a treatment vs. control study for small rivers since all segments within this size range had at least 10 headwater impoundments within their watersheds. Consequently, for cumulative effects studies on larger streams a correlative approach would have to be taken. In most instances the potential replicates in the largest size class provided a good range of values for the number of headwater impoundments within the watershed that was suited to a correlative study design.

There were surprisingly few differences in the replication potential between the cluster results generated by the Full and Reduced Sets. One major advantage of the Reduced Set is that it is easier to find clusters that are relatively "homogenous" in terms of landscape character, which increases the number of options available (i.e., stream types) for designing a study. We expected that the replication potential would dramatically decrease as we increased the number of clusters in both the Full and Reduced Sets. Yet, at least for the 12 stream types that we examined, there appears to be little change in replication potential. Overall, the number of potential replicates does show a decline as you increase the number of clusters, yet there are also instances where replication potential actually increases with the number of clusters.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Using GIS to Identify Potential Sample Sites for Assessing the Effects of PL-566 Structures on the Ecological Integrity of Missouri's Streams

Background

The Watershed Protection and Flood Prevention Act (PL-566) authorizes the USDA Natural Resources Conservation Service to help local organizations and units of government plan and implement watershed projects. PL-566 watershed projects are locally led to solve natural and human resource problems in watersheds up to 250,000 acres in size. These projects can include flood prevention and damage reduction, development of rural water supply sources, erosion and sediment control, fish and wildlife habitat enhancement, wetland creation and restoration, and increased recreational opportunities. Flood prevention measures include land treatment, and structural or non-structural measures that reduce or prevent floodwater damage by reducing runoff and erosion, or reducing the frequency or severity of flooding. Structural flood-control measures typically involve the construction of numerous headwater impoundments within the project watersheds.

The scientific literature suggests that flood control structures can have both positive and negative effects on biological integrity. In Missouri, there has been no explicit examination of the potential benefits or impacts of these small (<400 acres drainage area) flood control structures on biological integrity. The Clean Water Act mandates restoration, maintenance and protection of the physical, chemical and biological integrity of our nation's waters. Due to the potential effect of these structures on biological integrity it is essential that pertinent scientific data be compiled to specifically address these issues in Missouri. Such scientific information is critical to identify benefits and avoid or minimize impacts of these planned structures.

As a result of questions raised during the 401/404-permit process, as to the potential effect of these structures on biological integrity, an interagency advisory committee was formed to help devise strategies to answer many complex questions. Agencies represented on the committee include: Natural Resources Conservation Service (NRCS), Missouri Department of Conservation (MDC), Missouri Department of Natural Resources (MDNR), University of Missouri-Columbia (UMC), United States Geological Survey (USGS), U.S. Fish and Wildlife Service (USFWS) and U.S. Environmental Protection Agency (EPA). The committee devised the following goal, research questions, and objectives to address the issue discussed above.

Goal: Determine the positive or negative effects of headwater impoundments on hydrology, physical habitat, energy dynamics, water quality and biological interactions in headwater, midreach and mainstem streams.

The interagency advisory committee formulated the following questions, under each component of biological integrity, as an initial guide to devising a more detailed research strategy for addressing the above stated goal.

Hydrology:

1. Are there differences in the extent and the initiation of ephemeral, intermittent, and perennial flow during base-flow in project watersheds vs. non-project watersheds?
2. What is the history and the present spatial distribution of disturbance in project watersheds and how is it affected by the structures?

Physical Habitat:

1. How is habitat maintenance affected by altered hydrology, specifically sediment transport and flows, both locally and cumulatively?
2. Are there changes in riparian vegetation and riparian wetland habitat characteristics?
3. Are there differences in substrate composition and diversity, channel morphology, relative abundance of hydraulic habitat units (pools, riffles, runs) in project watersheds vs. non-project watersheds?

Energy Dynamics:

1. Are there differences in relative percentages of Coarse Particulate Organic Matter (CPOM), Fine Particulate Organic Matter (FPOM), and Dissolved Organic Carbon (DOC) in project watersheds vs. non-project watersheds?

Water Quality:

1. Are there local and cumulative effects on water quality both during construction and long-term?

Biological Interactions:

1. Are there local and cumulative effects to fish and invertebrate composition and diversity?
2. Are there effects to T&E species such as the Topeka Shiner?

Objectives

Prior to addressing the above or additional research questions, the committee determined that is was first necessary to assess existing approaches and research findings of projects that have examined the potential effects of impoundments on the biological integrity of streams.

<u>Objective 1:</u> Conduct a review of the existing scientific evidence regarding the influence of small impoundments on stream environments.

The committee further recognized the fact that northern Missouri, where most PL-566 structures have been constructed, contains a diversity of stream ecosystems. Since comparative or associative research designs might be used to address the above questions, the committee agreed that some effort must be made to ensure comparisons, of sites with and without a treatment (e.g., headwater impoundment), are only made among sites expected to have relatively similar ecological character in the absence of the treatment. Failure to do so could lead to “false positives” (treatment effect identified when one does not exist) or “false negatives” (no treatment effect identified when one does exist). To address this potential problem it was agreed that a GIS project should be undertaken to help identify potential study sites with similar watershed characteristics. This project would allow the committee to assess opportunities and limitations for addressing specific research questions when using either a comparative or correlative research design.

<u>Objective 2:</u> Use GIS to identify potential site replicates for assessing the effects of PL-566 structures on the ecological integrity of Missouri’s streams

Objective 1 was addressed in project conducted by the Missouri Cooperative Fish and Wildlife Research Unit, at the University of Missouri (Rabeni and Doisy 2004). Objective 2 was addressed by a project conducted by the Missouri Resource Assessment Partnership (MoRAP), which is the focus of the remainder of this report.

Study Area

The study area for Objective 2 encompasses approximately 46,000 square miles (Figure 1). Most of the study area falls within the Central Dissected Till Plains ecological subsection (Nigh and Schroeder 2002). Smaller components fall within the Osage Plains and Outer Ozark Border.

Methods

Several related tasks were required to achieve our overall objective for this project. Specifically, these tasks, in chronological order, included:

1. Select and/or create a digital stream network coverage that was suited to the overall objective.
2. Calculate various measures of stream size and gradient for each stream segment in the digital stream network.
3. Identify headwater impoundments.
4. Identify stream segments that intersect headwater impoundments.
5. Calculate the number of headwater impoundments within the watershed of every stream segment.
6. Classify stream segments into relatively distinct groupings based on watershed geology, soils, and landform.
7. Classify stream segments into relatively distinct groupings based on both watershed and local land cover.
8. Develop a GIS database that could be used to generate various experimental designs for assessing the potential positive or negative effects of headwater impoundments on northern Missouri streams.
9. Assess opportunities or limitations for treatment vs. control and correlative experimental designs for three stream size classes.

A general description of each of these tasks is provided below, starting on page 5.

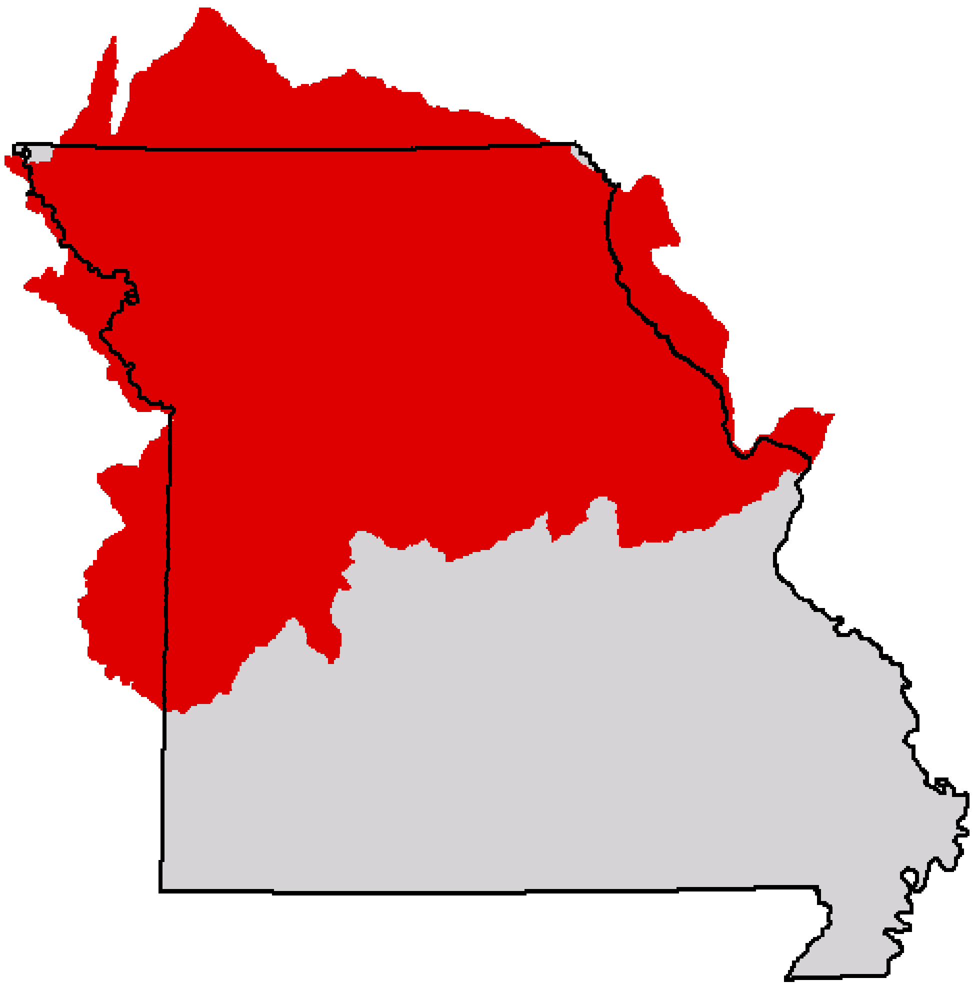


Figure 1. The study area (in red) for this project.

Task 1: Select and/or create a digital stream network suited to meet the overall objective

We originally intended to use the recently completed 1:24,000 National Hydrography Dataset (NHD) as the base data layer for our project. However, some of the Arc Marco Language (AML) programs needed to complete this project require a single-line, unidirectional, stream network. Unfortunately, the NHD contains numerous braided stream segments (Figure 2). Because of the immense about of manual labor and time required to remove these braids and identify primary path flows, we elected to create our own high-resolution stream network for this project.

We created our digital stream network using ESRI's ArcInfo Grid program. We first ran the FLOWDIRECTION grid command on the 30-meter National Elevation Dataset Digital Elevation Model (DEM) that encompassed our entire study area. FLOWDIRECTION creates a grid of flow direction from each cell to its steepest downslope neighbor (Greenlee 1987; Jenson and Domingue 1988). Using the resulting output grid, we then ran the FLOWACCUMULATION grid command. FLOWACCUMULATION creates a grid of accumulated flow to each cell by accumulating the weight for all cells flowing into each downslope cell (Jenson and Domingue 1988, Tarboton et. al 1991).

We then used the resulting output grid from the FLOWACCUMULATION command to create a digital stream network. This was accomplished by setting a threshold value for the number of cells required to initiate a stream channel. This threshold represents the total number of cells flowing into a single cell. As the threshold value decreases the density of the resulting digital network increases and the first order channels extend closer to the drainage divides.

To determine an appropriate threshold number for this specific project we used existing map data for the Big Creek-Hurricane Creek PL-566 project watershed, which showed the specific location of proposed headwater impoundments. First we created various digital stream networks using several threshold values (e.g., 50, 100, 200, 300 cells). We then scanned the maps of proposed headwater impoundments from the Big Creek-Hurricane Creek Proposal (USDA 1985) and digitally rectified these images. Our intent was to identify which threshold value produced a stream network that was dense enough and extended high enough into the drainage to touch each of the proposed headwater impoundments within the Big Creek-Hurricane Creek watershed. After presenting the preliminary results to, and consulting with, the advisory committee it was agreed that the 300-cell threshold produced the best results and thus this threshold was used to create the digital stream network for the entire study area. A 300-cell threshold translates to channels becoming initiated at a drainage area of 0.27 Km^2. For the entire study area, the resulting stream network contained 217,860 arcs (i.e., stream segments), of which 170,345 segments have their entire watershed area within the state of Missouri.

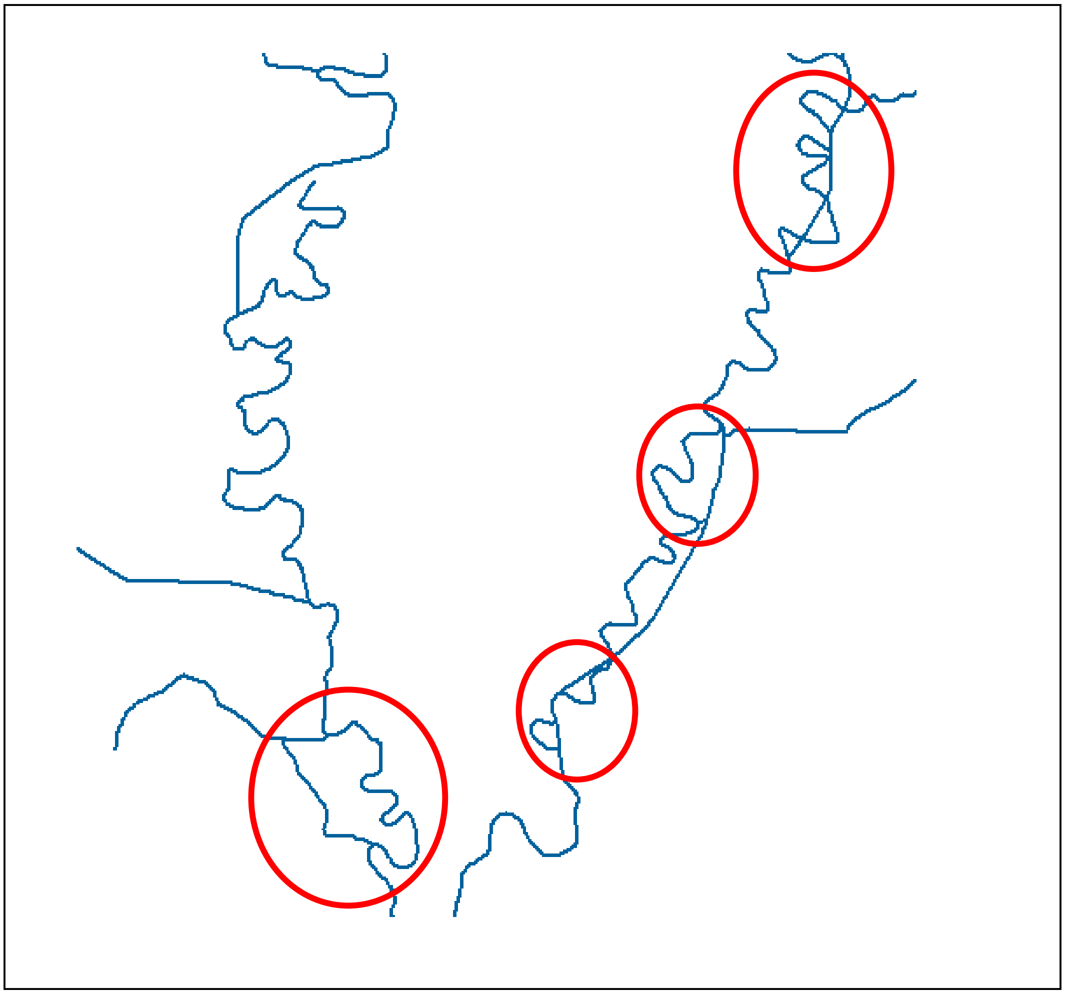


Figure 2. Illustration of the braids and loops within the 1:24,000 National Hydrography Dataset (NHD) that prevented us from using this dataset for the project.

Task 2: Calculate measures of stream size and local gradient for each segment

It has long been recognized that a wide array of structural features and functional processes, occurring within and along stream ecosystems, tend to change in a longitudinal continuum from the smallest headwaters to the largest rivers (Vannote et al. 1980). Consequently, studies designed to examine the potential influence of a given factor (other than drainage area) on the ecological character of streams, must somehow account for differences in stream size among potential study sites.

Instead of using the more precise measures of drainage area or discharge most investigators have utilized discrete stream size classes (Sensu Horton 1945 and Strahler 1957) in order to more tractably account for longitudinal changes in the abiotic and biotic character of streams. The Strahler ordering system is certainly the most widely recognized and the one most often used by stream ecologists for research and management (Hansen 2001). However, Strahler order often underestimates stream size due to vagaries in drainage network structure (Hynes 1970). With the Strahler ordering system it is common to have lower order streams (e.g., 3rd) with substantially larger drainage areas than higher order streams (e.g., 5th). Recognizing this problem Shreve (1966) devised another measure of stream size, termed link magnitude, which overcomes this problem since it is much more precisely related to drainage area (Hansen 2001). Link magnitude simply reflects the number of first order stream channels above a given stream segment.

Both of the above measures of stream size were calculated for each stream segment within the study area. We used the Stream_o.aml program, developed by the US Forest Services Redwood Sciences Laboratory (Lamphear and Lewis 1994), to compute the Strahler Order for each arc in the network. We then used the Shreve.aml program, which was originally developed by the Missouri Department of Conservation and subsequently modified to work with this project, for computing Shreve link magnitude for each arc. This AML utilizes the Arcplot command TRACEACCUMULATE to accumulate the number of streams with a Strahler stream order of 1 above each segment.

The specific drainage area above each stream segment was also calculated, however, the procedures for this are described below (Task 6 and 7). End users, therefore, have three options for grouping streams into various size categories; Strahler order, Shreve link magnitude, or drainage area. We recommend not using Strahler order since it provides a much less accurate depiction of stream size than the other two measures.

Stream gradient is another important variable to consider when devising any experimental design since it has long been recognized as a principle adjustable property of rivers that is often found to be associated with numerous abiotic and biotic factors within streams (Hack 1957; Knighton 1998; Nino 2002). Stream gradient was calculated for each individual stream segment in ArcView by using the same 30-meter DEM used to create the digital stream network. The minimum and maximum elevations were calculated for each segment. The minimum elevation was then subtracted from the maximum, divided by the stream length, and multiplied by 1000.

Task 3: Identifying headwater impoundments

Two geospatial datasets were evaluated for their ability to correctly identify existing headwater impoundments within the study area: 1) the Missouri's National Wetlands Inventory (NWI) and 2) MoRAP's 1992 land use land cover (LULC). We tested these two datasets within the Tabo Creek PL-566 Project Watershed (Figure 3). According to a 1999 Dams in Danger report (NRCS 1999) the Tabo Creek Watershed has 64 grade–stabilization dams. NRCS also provided us with digital point location data for each these PL-566 structures within the Tabo Creek watershed. These digital data corroborated the existence of 64 structures.

We then evaluated the ability of the two datasets (NWI vs. LULC) to identify the correct number of headwater impoundments within the Tabo Creek watershed and also assessed the amount of manual editing required by each dataset in order to achieve reasonable results. Each dataset was loaded into ArcView with the 300-cell digital stream network. Streams with a Strahler order of 1 or 2 (headwater streams) were then selected and intersected with the water bodies (no size restrictions on the size of the water body) in each of the two datasets. Based upon this subset of streams we identified 233 NWI water bodies and 122 LULC water bodies that intersected headwater streams in the Tabo Creek watershed (Figure 4). The most significant problem we encountered was with linear wetlands in the NWI data. Using existing attribution for the NWI data we were unable to select out only headwater water bodies since linear water bodies representing streams were always included in the selection. This problem with the NWI dataset could be fixed with extensive manual editing, but to do this for the entire study area was impossible given funding and time constraints.

Given the problems with the NWI data we proceeded with more tests of the LULC data. An obvious problem with our initial test was that we had placed no size restrictions on the water bodies within the LULC dataset. As a result, we were picking up larger impoundments and also many small "farm ponds" that were not the focus of the overall project. Using existing data on the size (i.e., areal extent) of PL-566 impoundments from NRCS (Elizabeth Cook, personal communication) we selected only those water bodies within the LULC dataset that were between 2 and 15 acres in size. We then intersected the first and second order streams with this subset of water bodies. This test correctly identified 61of the 64 existing PL-566 impoundments within the Tabo Creek watershed (Figure 5).

After presenting these test results to the advisory committee it was agreed that the MoRAP LULC dataset offered the best option for identifying headwater impoundments. It was also agreed that we should use two size restrictions for extracting water bodies from the LULC; 1) 2-15 acres and 2) 2-50 acres. The larger size range was included to provide some flexibility to the end user and also to account for potential classification errors in the LULC dataset. Consequently, we created two headwater impoundment

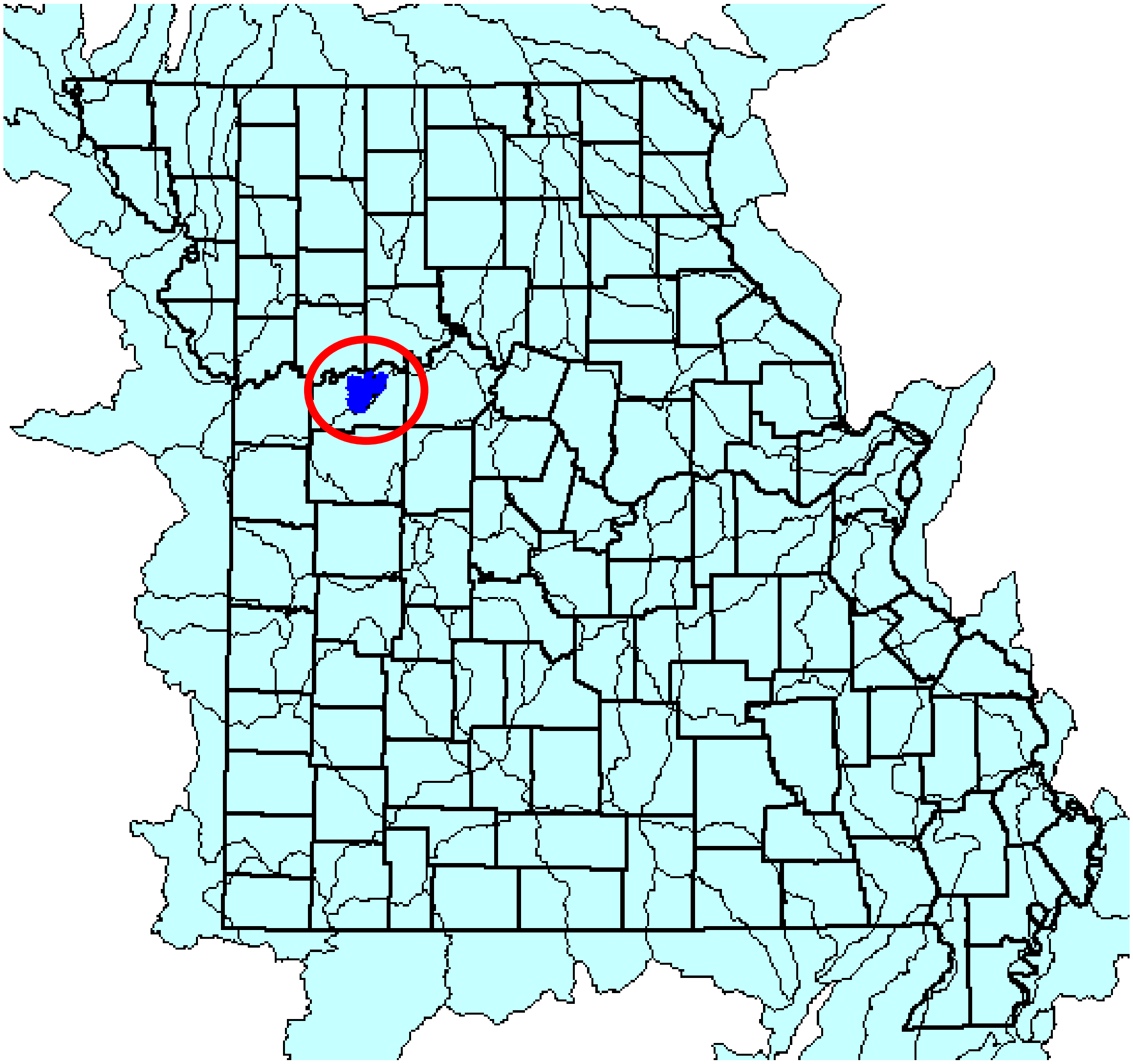
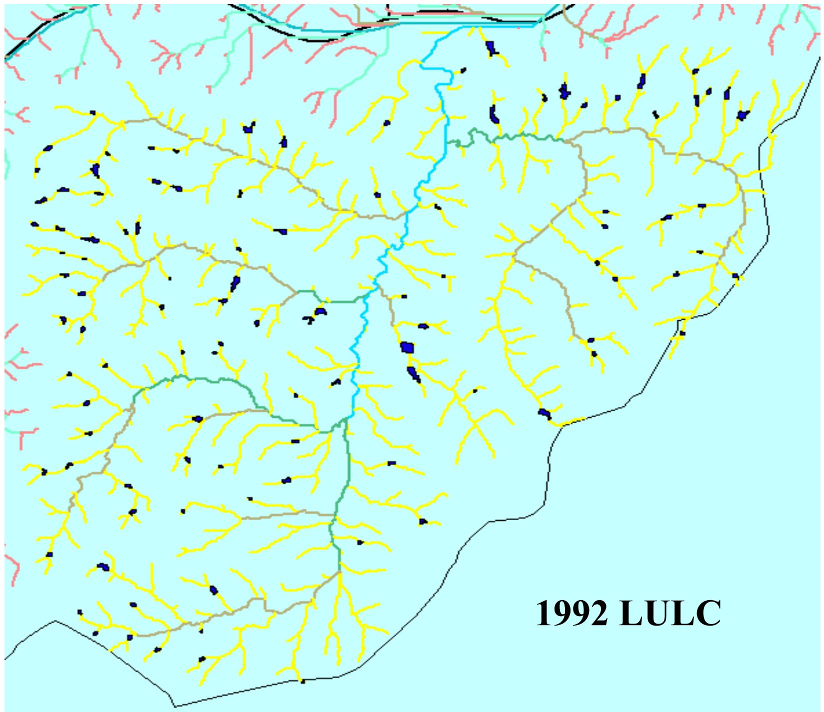


Figure 3. The Tabo Creek watershed that was used to assess differences in the ability of the NWI and the MoRAP LULC datasets to identify headwater impoundments.



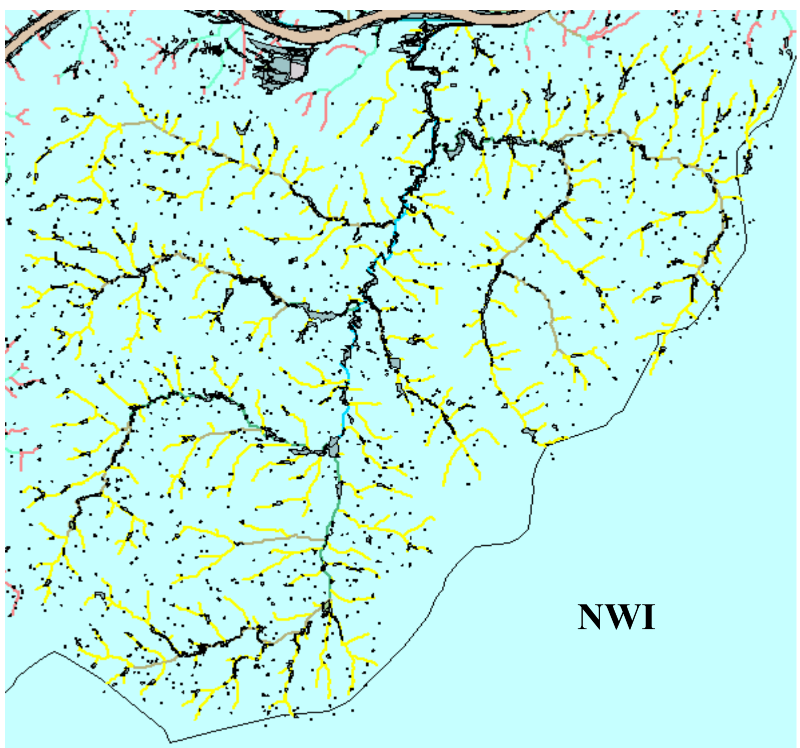


Figure 4. National Wetlands Inventory (NWI) and MoRAP 1992 Land Use Land Cover (LULC) waterbodies within Tabo Creek Watershed. Yellow streams are Strahler order 1 or 2.

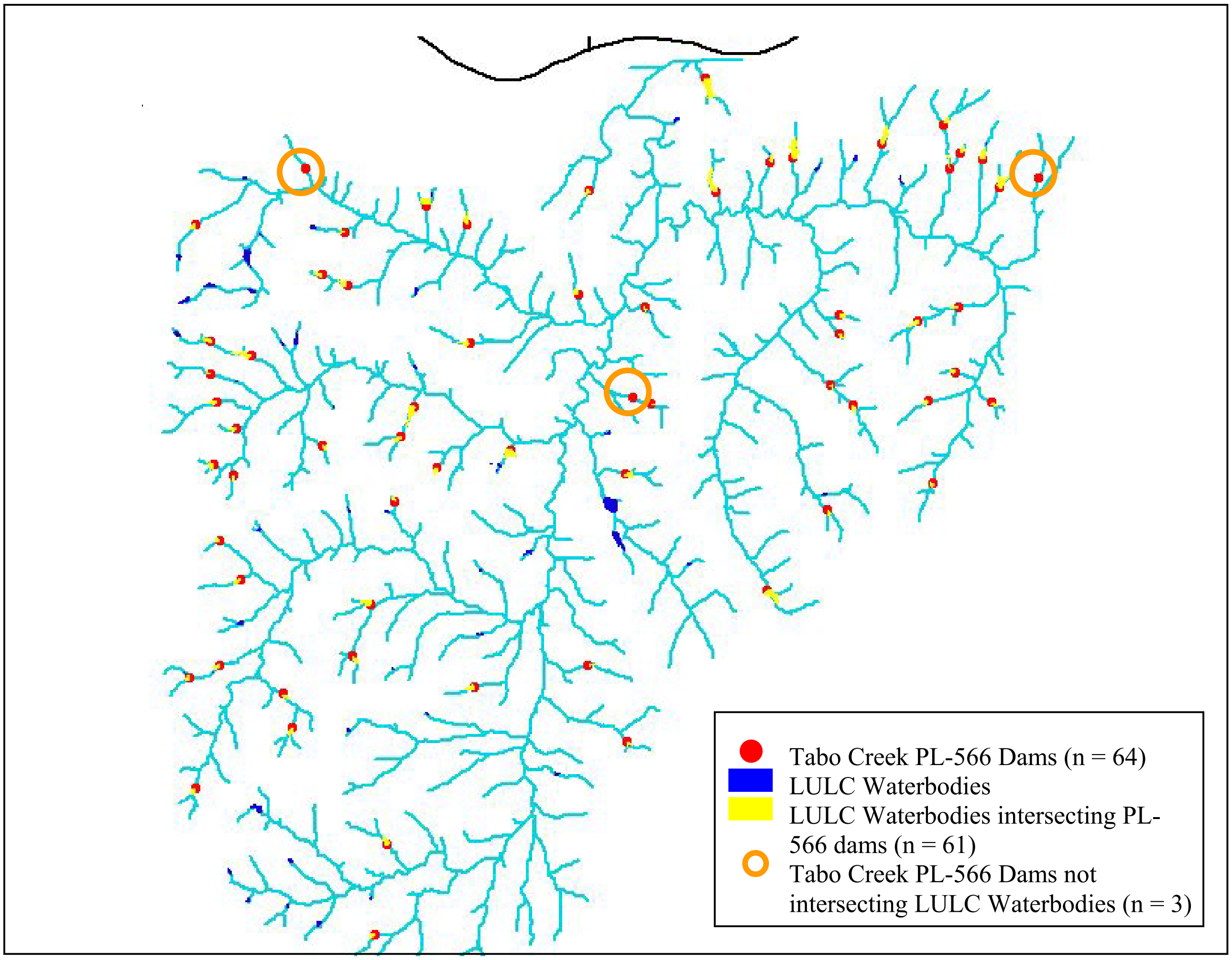


Figure 5. PL-566 structures within the Tabo Creek Watershed that were and were not captured by the MoRAP 1992 LULC.

coverages for the study area, one containing water bodies ranging from 2 to 15 acres in size and one containing water bodies ranging from 2 to 50 acres in size.

NOTE: An important caveat of using the MoRAP LULC data is that it is based upon 1991-93 satellite imagery. Impoundments constructed after this period are not captured in the resulting GIS datasets that will be used to help select potential study sites. Also, because we were unable to specifically identify PL-566 structures, studies specifically designed to assess these structures will have to rely on additional information sources to isolate these structures from other headwater impoundments.

Task 4: Identifying stream segments that intersect headwater impoundments

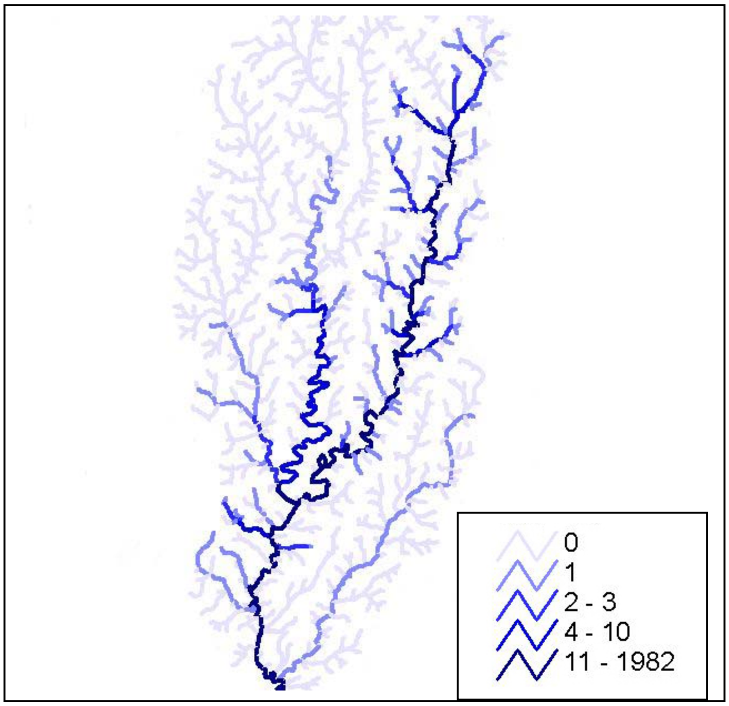
Once we had created the two water body coverages (i.e., 2-15 and 2-50 acre), from the LULC for the study area, we then proceeded to identify streams that intersected these water bodies. From the 300-cell digital stream network we selected all first and second order stream segments. We then intersected this subset of stream segments with the 2-15 acre water body coverage and also the 2-50 acre water body coverage. Separate attribute fields were created within the stream network coverage to hold this information. In each case, segments that intersected a water body were given an attribute value of 1, while all other segments were attributed with a 0.

During this attribution process we noticed numerous instances where stream segments that should have been attributed as intersecting a water body were not being appropriately attributed. Upon closer examination we found that in most instances this was the result of the fact that the 300-cell network was stopping just shy (i.e., tens of meters) of the outlet of the water body. To compensate for this problem we ran the above processes again using a 100-meter tolerance, as opposed to generating an entirely new digital stream network with a slightly smaller threshold value (e.g., 290 cells). This 100-meter tolerance allowed us to identify a stream as intersecting a water body if it was within 100 meters of that water body.

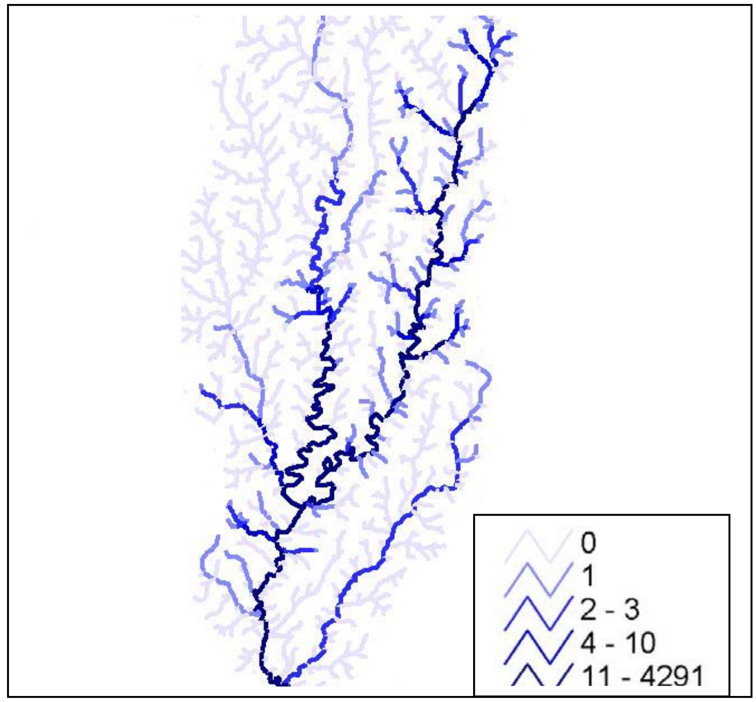
Consequently, the end user has four options for attempting to identify streams that intersect existing headwater impoundments; 1) water bodies 2-15 acres with no tolerance, 2) water bodies 2-50 acres with no tolerance, 3) water bodies 2-15 acres within a 100-meter tolerance, or 4) water bodies 2-50 acres within a 100-meter tolerance.

Task 5: Calculate number of headwater impoundments in the watershed of each segment

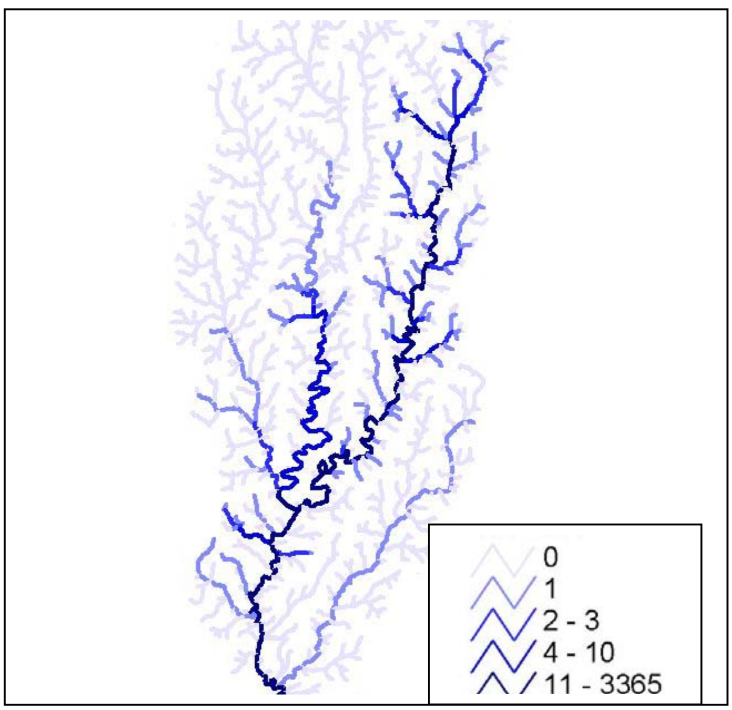
In order to design studies to examine the potential cumulative effects of multiple headwater impoundments on larger streams it was necessary to calculate the total number of headwater impoundments within the overall watershed of each stream segment. To accomplish this we used the TRACE ACCUMULATE command in ArcPlot to sum the total number of headwater impoundments above each segment using the four datasets generated in Task 4 (Figure 6).



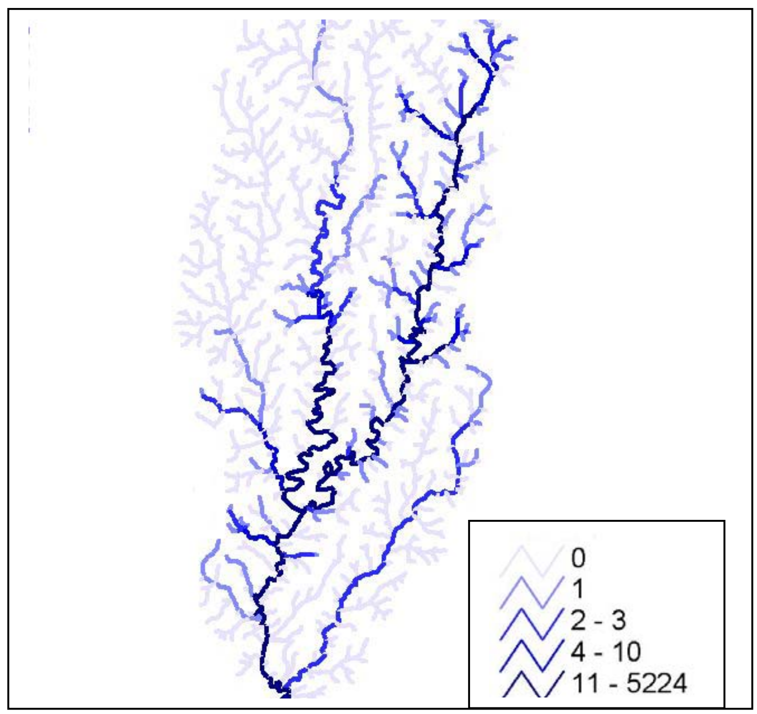
Accumulation of headwater ponds
2-15 acres



Accumulation of headwater ponds
2-15 acres with 100-meter tolerance



Accumulation of headwater ponds
2-50 acres



Accumulation of headwater ponds
2-50 acres with 100-meter tolerance

Figure 6. Accumulation of headwater ponds using the four different definitions.

Tasks 6 and 7: Classify stream segments into relatively distinct groupings

The ecological character of a particular stream segment is determined by a myriad of landscape features and associated processes operating at multiple spatiotemporal scales (Matthews 1998). Of particular interest are those features and processes operating within and immediately adjacent to the segment of interest and also those operating within the overall watershed (Lammert and Allan 1999; Wang et al. 2003). At the watershed-scale, geology, soils, landform, vegetation, and land use are the principle factors that collectively interact to determine a stream's ecological character (Hynes 1975; Panfil and Jacobson 2001).

Consequently, for tasks 6 and 7 we needed to generate watershed percentages for various landscape/land use features for of each of the 200,000 plus stream segments within the study area. To accomplish this we first used an AML program created by The Nature Conservancy's Freshwater Initiative (sheds.aml; TNC 2000). This AML uses a DEM and digital stream network coverage to generate polygons that represent the immediate drainage of each stream segment, which we call segmentsheds (Figure 7). The AML also adds an item, ARCIDNUM, which relates the resultant segmentshed polygons back to the appropriate arcs in the stream network.

One limitation of the sheds.aml is that it can only process 100,000 arcs at a time. As a result, we had to divide the stream network for the entire study area into 3 separate coverages: streamnet1 (Missouri River east), streamnet23 (Missouri River west/Osage River), streamnet4 (Mississippi River). We ran the sheds.aml on each of these coverages, which resulted in 3 corresponding segmentshed coverages: sheds1, sheds23, sheds4.

For each of the landscape features included in Tasks 6 and 7 (i.e., geology, soils, relief, and land cover) the same data processing steps were taken. Data was loaded into ArcView and the actual and percent area of each feature class (e.g., land cover classes; water, urban, grassland, cropland, forest) was calculated for each individual segmentshed polygon. These data were then transferred from the segmentshed polygon coverages (sheds1, sheds23, sheds4) to the stream network coverages (streamnet1, streamnet23, streamnet4) using the common item ARCIDNUM. The TRACE ACCUMULATE command was then used to summarize the overall and percent area of each feature class within the entire watershed of each stream segment (Figure 8). Finally, once the processing for the three separate coverages was completed, they were merged back into a single coverage for the entire study area.

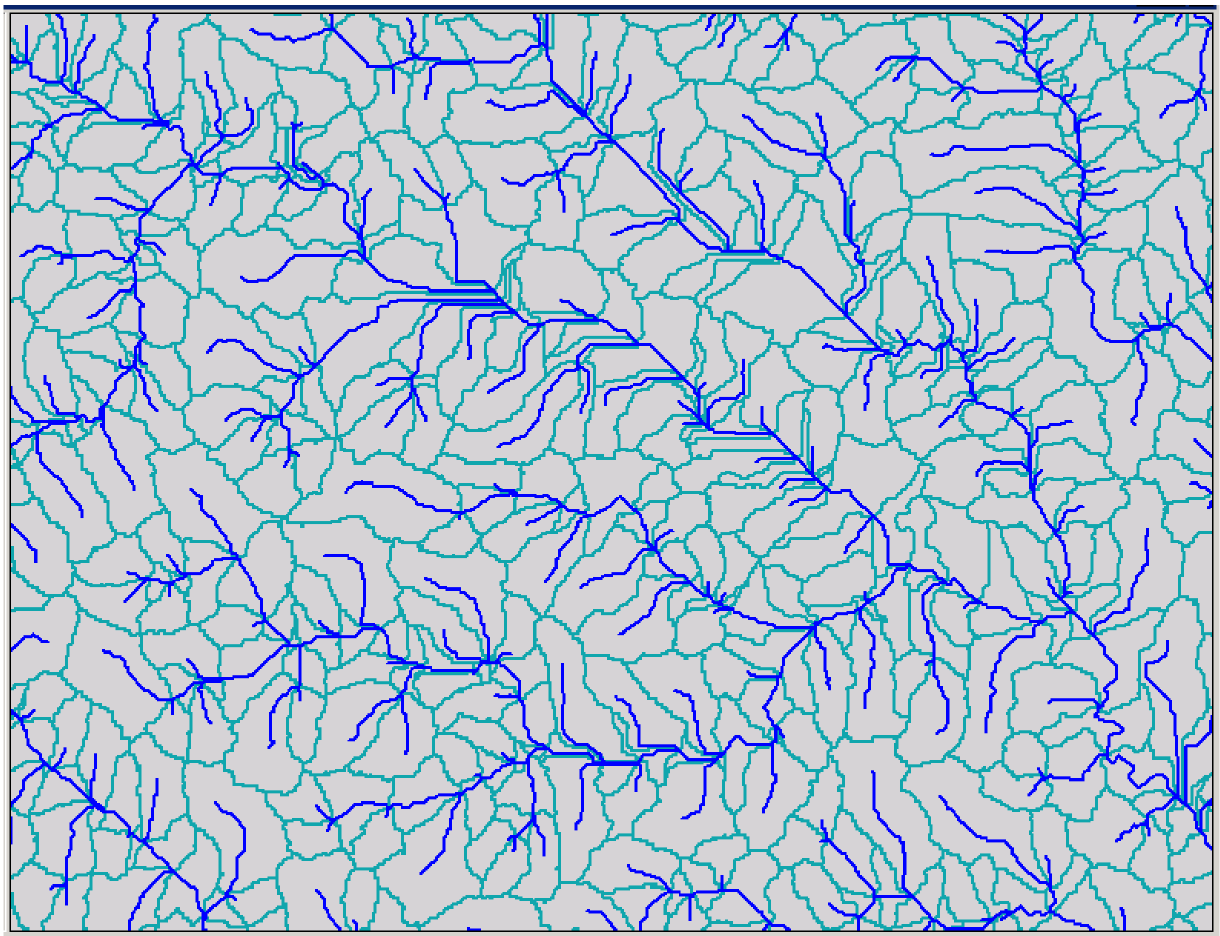


Figure 7. Map showing segmentsheds (immediate drainage) for each individual stream segments.

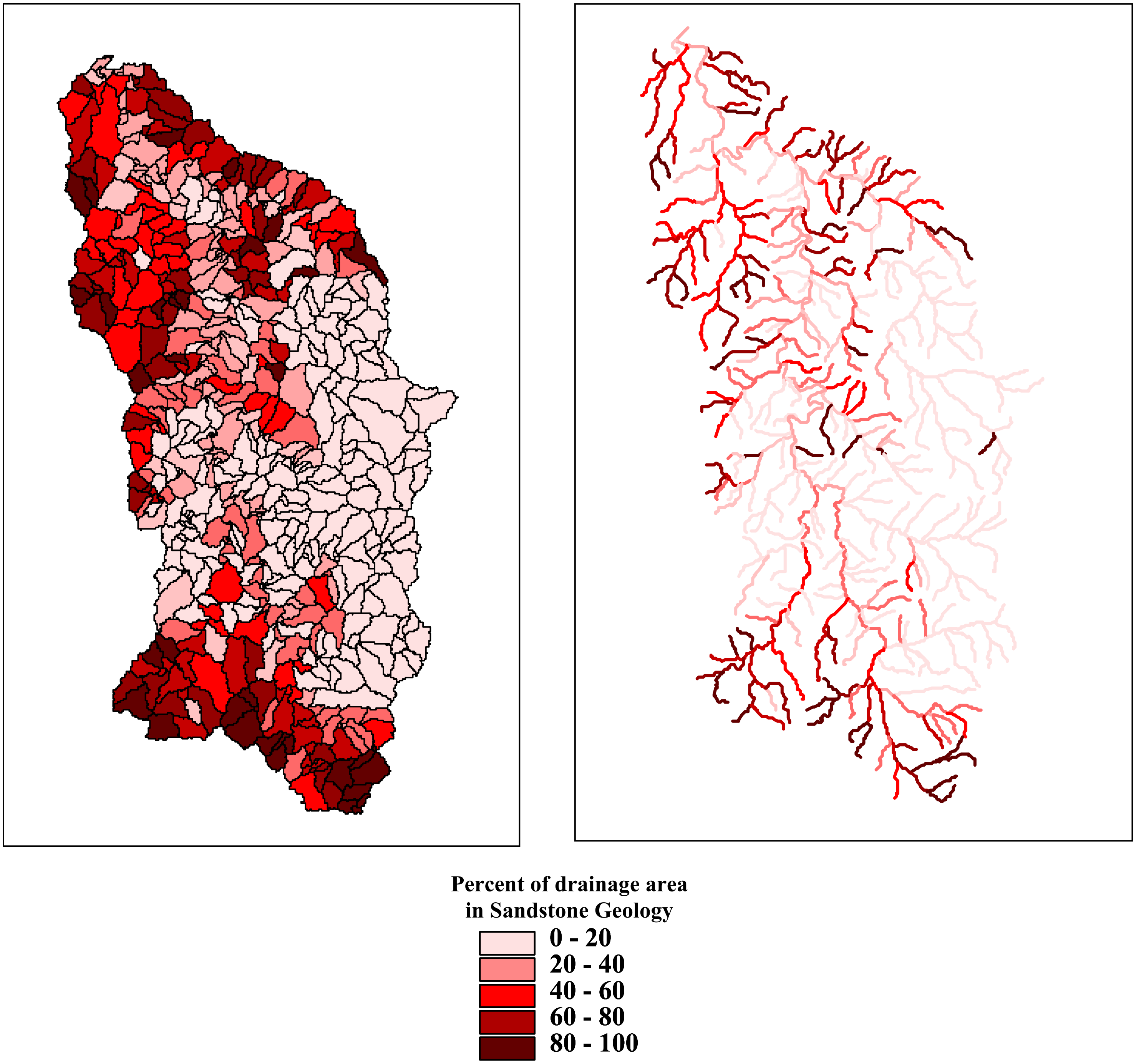


Figure 8. Map example showing how the TRACE ACCUMULATE command can be used to generate overall watershed percentages for each individual stream segment. This example shows the percent of sandstone geology within each segments watershed. The data can be captured and visually represented by either the segmentsheds or the corresponding stream segments.

Task 6 and 7 Landscape/Land Cover Datasets

<u>Geology</u>

We used the 1:500,000 statewide digital geology data for Missouri created in 1992 by digitizing from the 1979 Geologic Map of Missouri (Figure 9; MDNR 1992). From this coverage we calculated the area and percent area for the System and Series-level classes and also the Gentype classes from the Missouri Aquatic GAP Project for each segmentshed polygon and each segment's overall watershed (Table 1). Within our study area, Series-level geology provides the finest breakdown with 12 different classes, System-level geology has 6 classes, and Gentype has only 5 classes. After conferring with the advisory committee, it was decided that System geology should be used in the subsequent cluster analyses, discussed below.

<u>Soils</u>

Our original intention was the use the higher resolution Missouri SSURGO II data from the NRCS. We gathered data for the 68 Missouri counties that covered the study area and began extracting information on Available Water Capacity, Permeability, and Hydrologic Soil Group. However, with the level of detail in the data and the number of segmentsheds in our dataset, we reached a limit of the GIS software. To process these data for the entire study area would have required breaking the area into numerous subsets, processing each subset separately, and then merging everything back together after each subset was completed. Funding and time constraints prevented us from using this approach.

In lieu of SSURGO data we decided to use STATSGO data (Figure 9). From the STATSGO coverage we calculated the area and percent area of each Hydrologic Soil Group and Surface Texture class for each segmentshed and segment's overall watershed (Table 2). We also created a third soils variable by condensing the original 12 Surface Texture classes into 5 general classes (Table 2). For statistical reasons, we used the 5 general classes of Surface Texture in the cluster analyses.

<u>Relief</u>

To characterize the landform of each segment's watershed we first created a relief grid for the study area by using the grid command FOCALRANGE. For each cell in the input grid, this command finds the range of the values (maximum and minimum) within a specified neighborhood and sends it to the corresponding cell location on the output grid. We used a 1-Km² circle to define the neighborhood. The minimum values were then subtracted from the maximum values to generate a relief value for each cell.

The resulting relief grid ranged from 0-650 feet for the study area. This range was then broken into 6 relief classes (0-50, 51-100, 101-200, 201-300, 301-500, 501-650) based upon the divisions used to create the Missouri Land Type Associations (Figure 9; Nigh and Schroeder 2002). We then calculated the area and percent of each relief class within each segmentshed polygon and each segment's overall watershed.

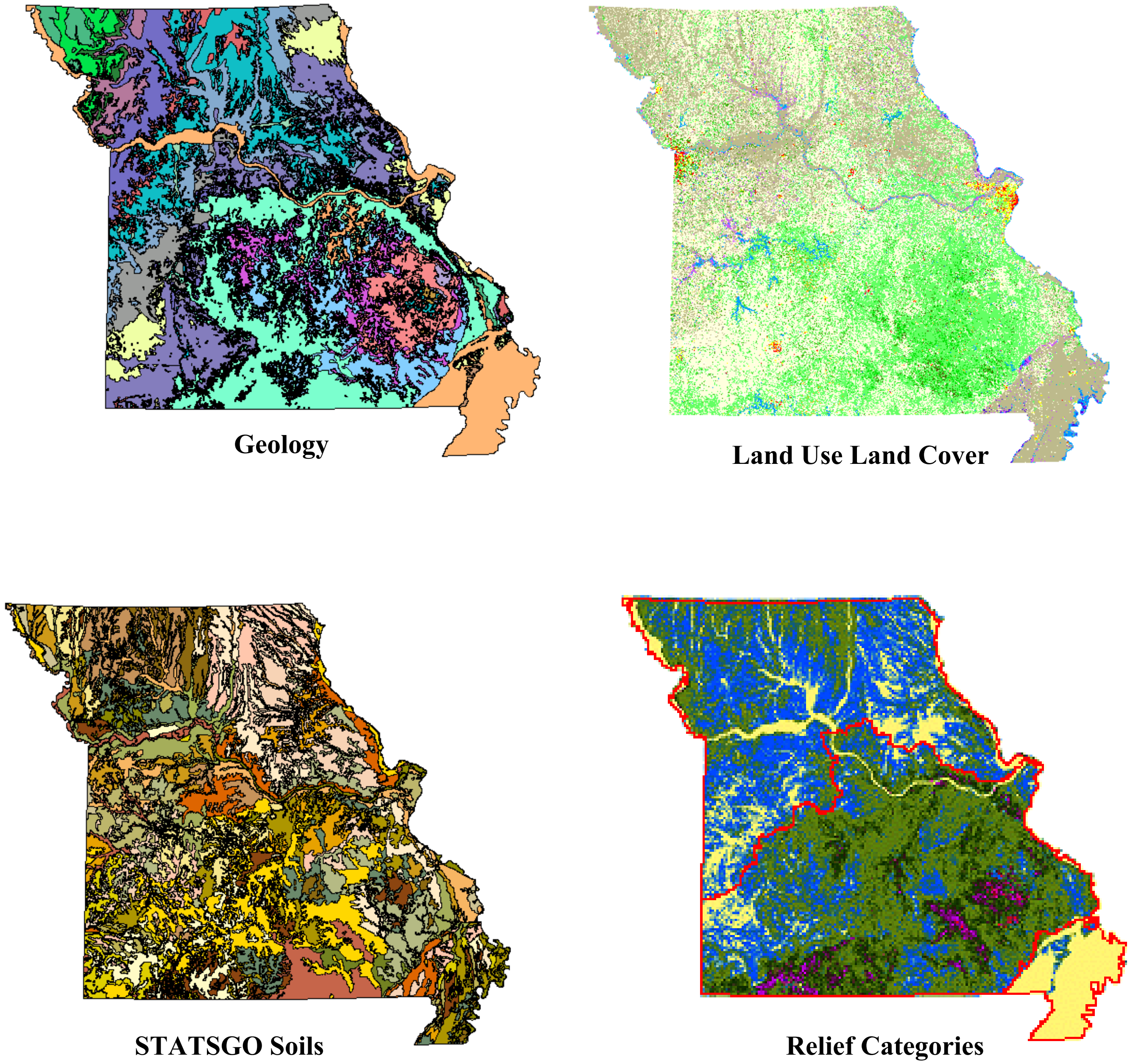


Figure 9. Maps showing the geospatial datasets used to generate the various landscape/land cover statistics for each stream segment, which were ultimately used in the cluster analyses to classify stream segments into distinct groups.

Table 1. Series, System, and General geology categories for which watershed percentages were generated for each stream segment in the study area.

Series	System	General Geology
Atokan/Desmoinesia	Devonian	Alluvium
Holocene	Mississippian	Clay
Virgilian	Ordovician	Dolomite
Desmoinesia	Pennsylvanian	Limestone
Missourian	Quaternary	Sandstone
Osagean	Silurian	
Alexandrian/Niagar		
Canadian		
Champlanian		
Kinderhookian		
Lower/Middle/Upper		
Meramecian		

Table 2. Hydrologic soil group and soil surface texture categories for which watershed percentage statistics were generated for each stream segment in the study area.

Hydrologic Soil Group	Surface Texture	Condensed Surface Texture
Hydrologic Soil Group B	Cherty/Silty Loam (CRSIL)	Cherty
Hydrologic Soil Group C	Very Cherty/Silty Loam (CRVSL)	
Hydrologic Soil Group D	Clay Loam (CL)	Clay
	Silty Clay (SIC)	
	Silty Clay Loam (SICL)	
	Fine Sandy Loam (FSL)	Sandy
	Loamy sand (LS)	
	Silty Loam (SIL)	Laomy
	Loam (L)	
	Variable (VAR)	
	Stony Laom (STL)	Stony
	Stony Silt Loam (STSIL)	

Land Cover

We used Missouri's 1992 Land Use Land Cover (LULC) dataset created at MoRAP for use in the Missouri Gap Analysis Project (Figure 9). The LULC dataset is available in 44, 16, and 6 classes. After consulting with the advisory committee it was decided that we should use the 6-class dataset that breaks land cover into urban, cropland, forest, grassland, swamp, and open water. It was also decided, due to classification problems, that the swamp class should not be included in the calculations. Consequently, we ended up calculating the area and percent area of urban, cropland, forest, grassland, and open water within each segmentshed polygon and each segment's overall watershed.

Task 6 and 7 Statistical Methods

Cluster analysis was used to identify groups of stream segments that are relatively similar with regard to watershed landscape character and also both watershed and local land use. Cluster analysis is a multivariate analysis technique that seeks to organize information about variables so that relatively homogeneous groups, or "clusters," can be formed. The resulting clusters should be internally homogenous (members are similar to one another) and externally heterogeneous (members within one cluster are *not* like members of other clusters).

Due to the extremely high number of stream segments (i.e., records) in the digital stream network generated for this project (over 175,000), we had to use the FASTCLUS procedure within SAS to perform all cluster analyses since either file or computing limitations of SAS or other software prevented their use (SAS 2001). The FASTCLUS procedure is specifically designed for clustering of very large data sets and can find good clusters with only two or three passes over the data (SAS 2001). It performs a disjoint cluster analysis on the basis of Euclidean distances computed from one or more quantitative variables. It is not a hierarchical clustering algorithm and therefore separate analyses must be performed for each of the desired number of clusters. FASTCLUS combines an effective method for finding initial clusters with a standard iterative algorithm for minimizing the sum of the squared distances from the cluster means. Specifically, the procedure first selects a set of points called "cluster seeds" as a first guess of the cluster means. Each observation is then assigned to the nearest seed to form temporary clusters. The initial seeds are then replaced by the means of the temporary clusters, and the process is repeated until no further changes occur within the clusters. Anderberg (1973) described this method as nearest centroid sorting.

Cluster analysis methods will always produce groupings, which may or may not prove useful for classifying objects of interest. If the groupings discriminate between variables not used to do the grouping (e.g., instream habitat) and those discriminations are useful, then cluster analysis is useful. Consequently, an assumption of our project is that the variables used to identify clusters (geology, soils, landform, and land use) are significantly related to the structure and function of the stream ecosystems. With this assumption we expect streams of similar size and also watershed geology, soils, landform, and land use to be similar with regards to water chemistry, energy dynamics, instream habitat, flow regimes, and resident biota.

Task 6 and 7 Input Datasets for Cluster Analyses

Based on input from the committee we ran cluster analyses on four separate datasets.

<u>Watershed Full Set</u>

The Watershed Full Set contained **watershed** percentage statistics for 20 total variables (6 geology, 6 relief, 5 soil texture, and 3 hydrologic soil groups) (Table 3).

<u>Watershed Reduced Set</u>

The Watershed Reduced Set contained **watershed** percentage statistics for 9 variables (6 relief and 3 hydrologic soil groups) (Table 4).

<u>Watershed Land Cover Set</u>

The Watershed Land Cover Set contained **watershed** percentage statistics for 5 variables (i.e., urban, grassland, cropland, forest, and open water).

<u>Local Land Cover Set</u>

The Local Land Cover Set contained **segmentshed** percentage statistics for 5 variables (i.e., urban, grassland, cropland, forest, and open water).

We ran separate cluster analyses for the land use/cover because we were first and foremost interested in the inherent natural differences among potential study sites and secondarily differences in existing land use. We were concerned that including the land cover/use with the other landscape variables could have obscured the results in some situations since variation in land use may override variation in the other landscape features. If land use was perfectly correlated with the other factors this would not have been a problem. However, if land uses such as cropland occur primarily in landscapes with certain combinations of geology, soil, and relief (e.g., optimal for crop production) then those sites that had land cover/use percentages that were anomalies because they occurred within landscapes with other combinations of landscape features (e.g., marginal for crop production) may have been inappropriately classified simply because variation in land use overrode differences in the other features.

Since all of the values for each variable were already relativized values (i.e., percentages) no transformations were performed on the data and the cluster analyses were run on the raw percentages. For each of the input datasets we generated 2 to 50 clusters, increasing by increments of 2 until we reached 20 clusters, at which point we began increasing the number by increments of 10 (Figure 10).

Table 3. Landscape variables and associated categories that were used in the Watershed Full Set cluster analyses.

Data Source:	1:500,000 Statewide Geology	STATSGO Soils Data		Digital Elevation Model
Data Type:	System Geology	Hydrologic Soil Group	Soil Surface Texture	Relief Category
Attribute Categories:	Devonian	Hydrologic Soil Group B	Cherty (CRSIL, CRVSL)	Relief Category 1 (0-50 feet)
	Mississippian	Hydrologic Soil Group C	Clay (CL, SIC, SICL)	Relief Category 2 (101-200 feet)
	Ordovician	Hydrologic Soil Group D	Loamy (L, SIL, VAR)	Relief Category 3 (201-300 feet)
	Pennsylvanian		Sandy (FSL, LS)	Relief Category 4 (301-500 feet)
	Quaternary		Stony (STL, STSIL)	Relief Category 5 (501-650 feet)
	Silurian			

Table 4. Landscape variables and associated categories that were used in the Watershed Reduced Set cluster analyses.

Data Source:	STATSGO Soils Data	Digital Elevation Model
Data Type:	Hydrologic Soil Group	Relief Category
Attribute Categories:	Hydrologic Soil Group B	Relief Category 1 (0-50 feet)
	Hydrologic Soil Group C	Relief Category 2 (101-200 feet)
	Hydrologic Soil Group D	Relief Category 3 (201-300 feet)
		Relief Category 4 (301-500 feet)
		Relief Category 5 (501-650 feet)

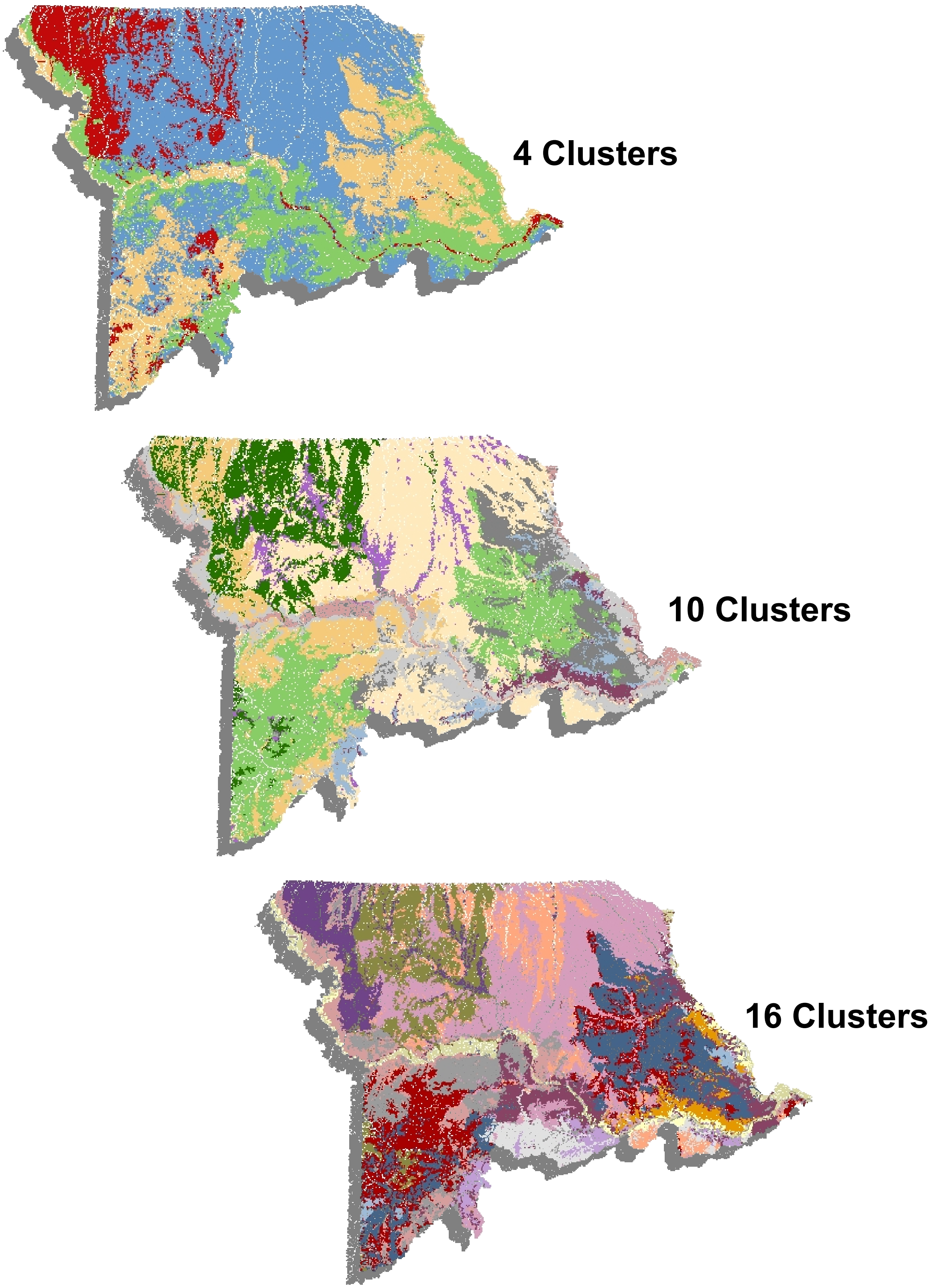


Figure 10. Maps depicting cluster analysis results (4, 10, and 16 clusters) generated using the Full Set of watershed landscape variables. These maps reveal that the cluster analysis initially groups sites according hydrologic soil groups (4 clusters) and then as the number of clusters increases relief, soil surface texture, and geology become progressively important in discriminating among clusters.

Task 6 and 7 Identifying the appropriate number of clusters

There are no completely satisfactory methods for determining the true number of clusters for any type of cluster analysis (Everitt 1979; Bock 1985; Hartigan 1985). Ordinary significance tests, such as analysis-of-variance F-tests, are not valid for testing differences between clusters. Since clustering methods attempt to maximize the separation between clusters, the assumptions of the usual significance tests, parametric or nonparametric, are drastically violated. For example, if you take a sample of 100 or 1000 observations from a single univariate normal distribution, have PROC FASTCLUS divide it into two clusters, and perform a t-test to compare the cluster means, you usually obtain a significant P-value (SAS 2001).

There are, however, various external or internal criteria that can be used to help determine the appropriate number of clusters within a particular multivariate data set (Jongman et al. 1995). External criteria are not dependent upon the method of clustering since independent data are used to test whether or not the clustering results are meaningful. However, in our case, external data, such as species composition or abundance, water chemistry, flow regimes, or instream habitat, were not available and therefore could not be used to assess the proper number of clusters. Internal criteria are dependent upon the data used for obtaining the clusters and also the specific clustering method. Most often two types of internal criteria are used to determine the optimum solution (Jongman et al. 1995). The first is the homogeneity of the clusters, which requires some measure of the (dis)similarity of the members of each cluster. The second is the degree of separation of the clusters, which requires some measure of the (dis)similarity of each cluster to its nearest neighbor. Typically, plots of these internal criteria against the number of clusters are used to guide the decision of how many clusters is optimal (Jongman et al. 1995; Salvador and Chan 2003).

In addition, PROC FASTCLUS provides estimates of the overall r-square, a pseudo F-statistic, and the cubic clustering criterion (Calinski and Harabasz 1974; Sarle 1983). Plotting these criteria against the number of clusters and then determining where these three criteria are simultaneously maximized also provides a good indication of the proper number of clusters within the overall dataset (Milligan and Cooper 1985; SAS 2001). However, caution must be used with these criteria when the discriminatory variables are correlated, which does occur in our case. It must also be emphasized that these criteria are appropriate only for compact or slightly elongated clusters, preferably clusters that are roughly multivariate normal.

We used all three of the internal criteria described above to provide insight into the proper number of clusters for each dataset. Specifically, we generated three separate diagnostic plots for each dataset.

1. Plots of the mean distance among cluster centroids versus the number of clusters. (Provides a means of assessing the degree of separation among clusters as the total number of clusters changes)

2. Plots of the mean root-mean-square distance between observations within clusters versus the number of clusters. (Provides a means of assessing the relative homogeneity of observations within clusters as the number of clusters changes).

3. Overlay plots of the overall r-square, cubic clustering criterion (CCC), and pseudo F-statistic values versus the number of clusters. (Provides a means of collectively assessing how much of the overall variance in the dataset is explained by the clusters (overall r-square), the significance/validity of the clusters against the null hypothesis of a multivariate uniform distribution (CCC), and relative significance of the differences among the cluster means (pseudo F-statistic) as the number of clusters changes).

Agreement among these diagnostic plots, as to how many clusters actually exist within the dataset, generally provides a good indication of the ideal number of clusters (Cooper and Milligan 1984; Milligan and Cooper 1985).

Task 6 and 7 Results: How many distinct clusters in each of the input datasets?

<u>Full Set of Variables</u>

- Diagnostic plots suggest that there are anywhere from 12 to 20 distinct clusters in the dataset. Likely, 14 to 18 clusters are optimal (Figures 11, 12, 13).

<u>Reduced Set of Variables</u>

- Diagnostic plots suggest that there are anywhere from 12 to 20 distinct clusters in the dataset. Likely, 12 to 16 clusters are optimal (Figures 14, 15, 16).

<u>Watershed Land Cover</u>

- Diagnostic plots suggest anywhere from 8 to 16 distinct clusters in the watershed land cover dataset. Closer examination of the cluster means suggests that at least 10 clusters are necessary to separate out the major land cover classes and their combinations, as well as those units that are dominated by the much less common water class. In other words, using fewer than 10 clusters, segments that have a high percentage of water within their watersheds would not be distinguished because they are essentially "hidden" within one of the other clusters. Assuming these segments are not desirable for addressing the goal of the broader project (since these are mainly segments currently impounded by larger reservoirs) it is important that enough clusters are used in order to isolate these segments and this occurs at or above 10 clusters (Figures 17, 18, 19).

<u>Local Land Cover</u>

- Plots suggest that there are anywhere from 12 to 20 distinct clusters in the dataset. As was found with the watershed land cover, at least 10 clusters is required in order to separate out those segments that fall within large water bodies (Figures 20, 21, 22).

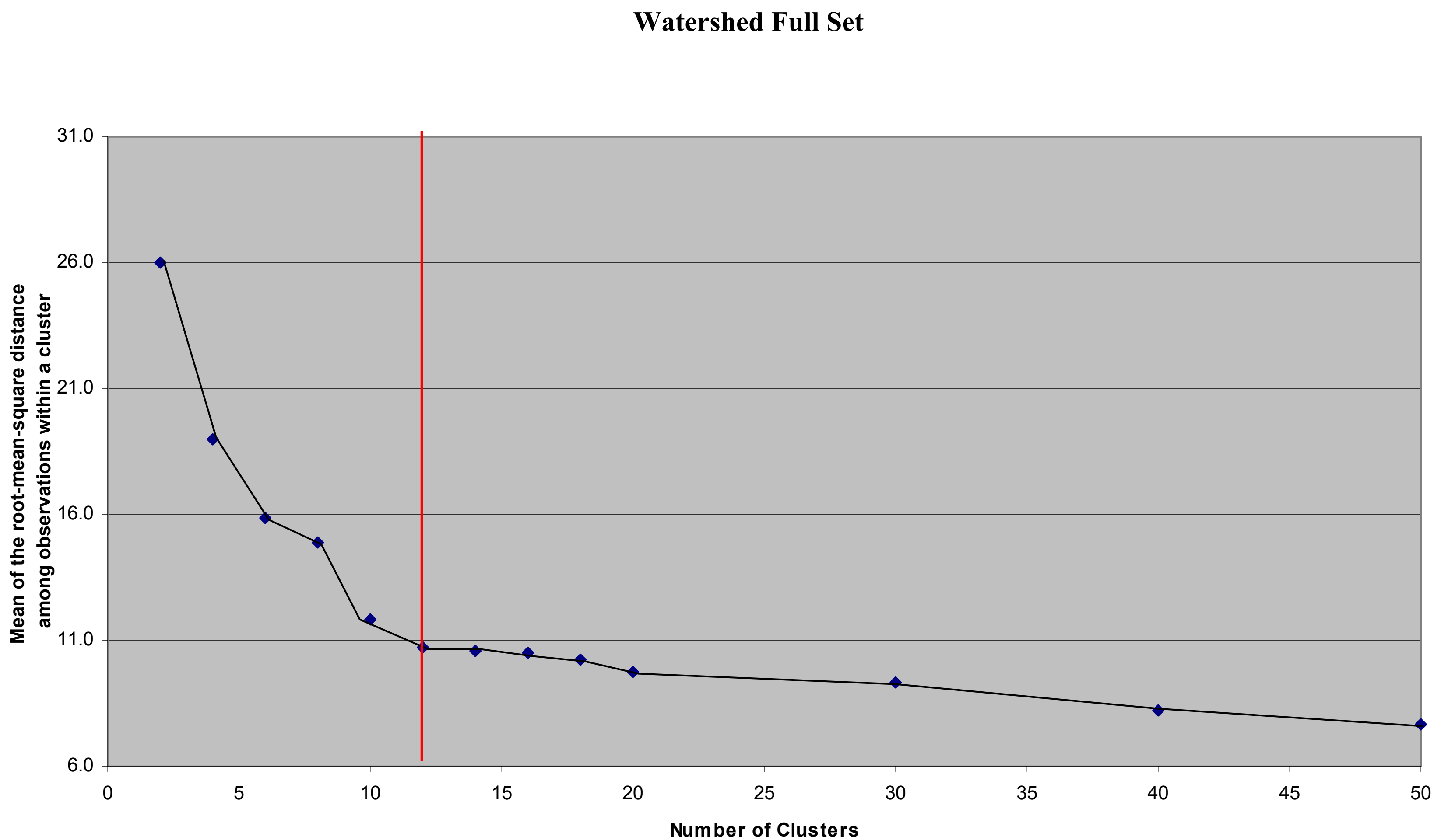


Figure 11. Line plot of the mean of the root-mean-square distance among observations within all clusters versus the number of clusters. This plot suggests that above twelve clusters only minimal additional variation, in the variables used to form the clusters (full set of landscape variables), is accounted for.

Figure 12. Line plot of the mean distance between cluster centroids versus the number of clusters. This plot suggests that above twenty clusters only minimal additional variation, in the variables used to form the clusters (full set of landscape variables), is accounted for.

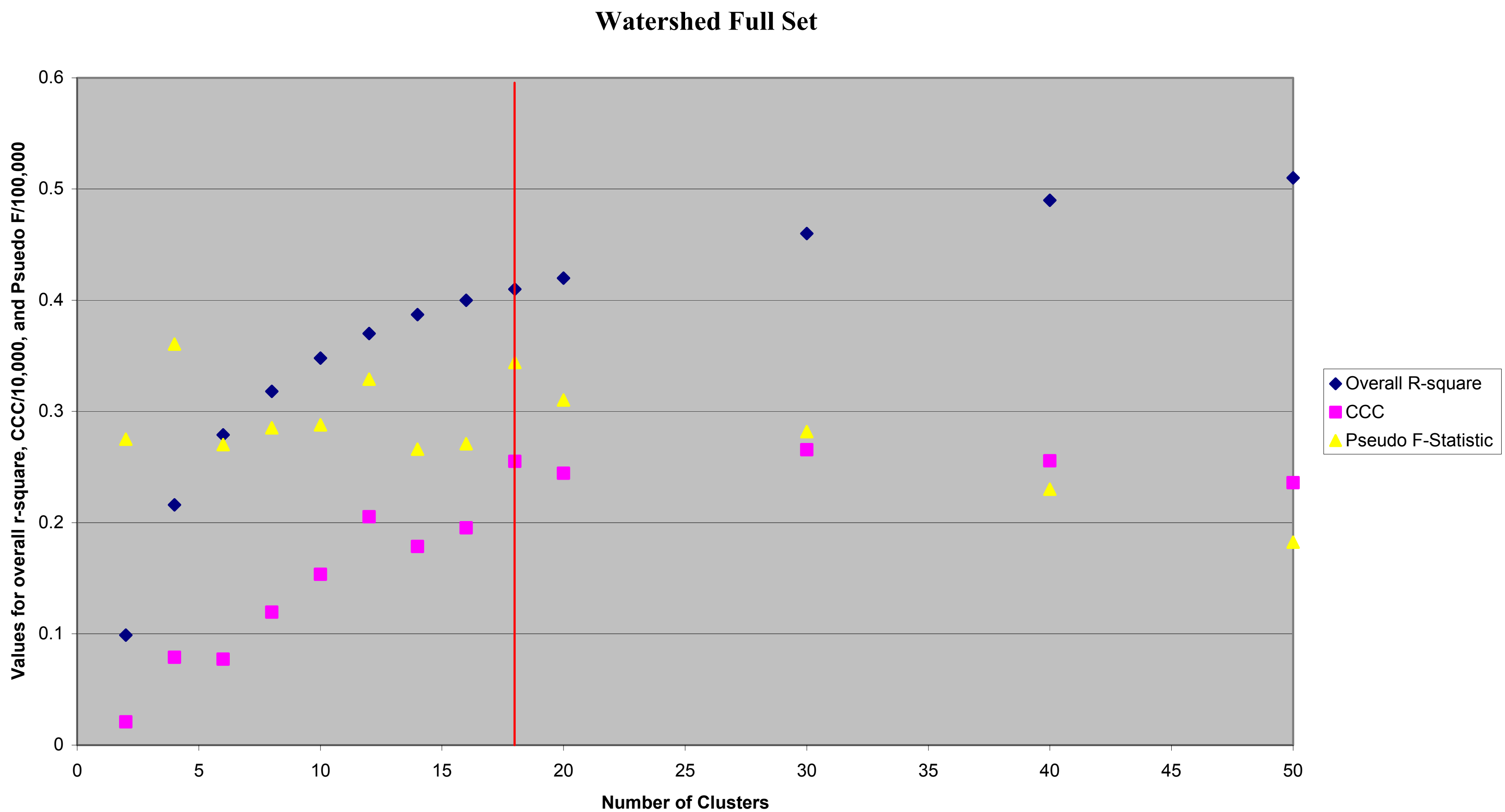


Figure 13. Plot of the overall r-square, cubic clustering criterion (CCC), and pseudo F-statistic values versus the number of clusters. This plot suggests that approximately 18 clusters is optimal for identifying relatively homogenous groupings for the overall set of landscape variables (system geology, soil surface texture, hydrologic soil group, and six relief classes). Note: for presentation purposes, the CCC and Psuedo F-statistic values were divided by 10,000 and 100,000, respectively.

Watershed Reduced Set

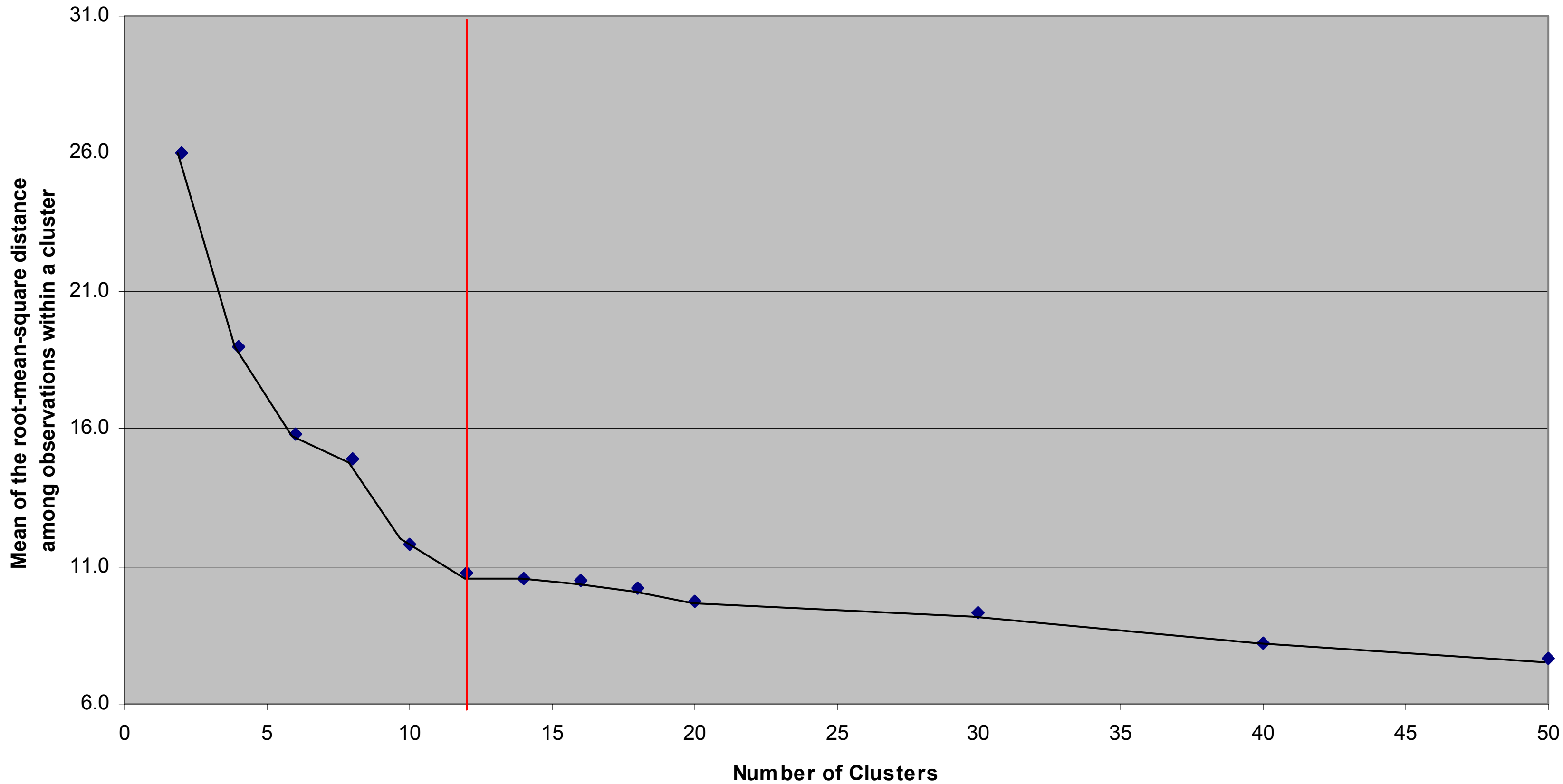


Figure 14. Line plot of the mean of the root-mean-square distance among observations within all clusters versus the number of clusters. This plot suggests that above twelve clusters only minimal additional variation, in the variables used to form the clusters (reduced set of landscape variables), is accounted for.

Figure 15. Line plot of the mean distance between cluster centroids versus the number of clusters. This plot suggests that above twenty clusters only minimal additional variation, in the variables used to form the clusters (reduced set of landscape variables), is accounted for.

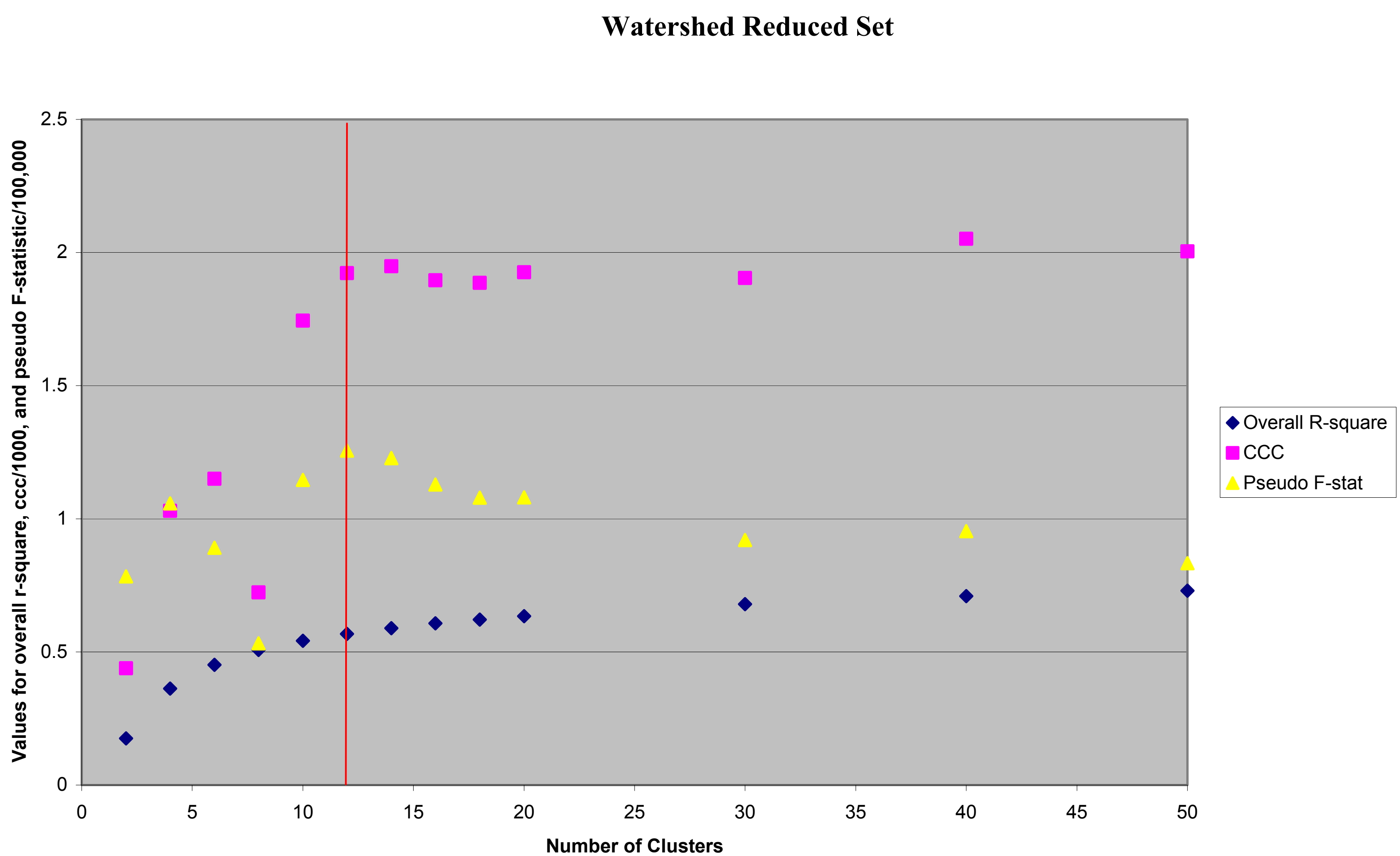


Figure 16. Plot of the overall r-square, cubic clustering criterion (CCC), and pseudo F-statistic values versus the number of clusters. This plot suggests that approximately 12 clusters is optimal for identifying relatively homogenous groupings for the reduced set of landscape features (hydrologic soil groups and six relief classes). Note: for presentation purposes, the CCC and Psuedo F-statistic values were divided by 10,000 and 100,000, respectively.

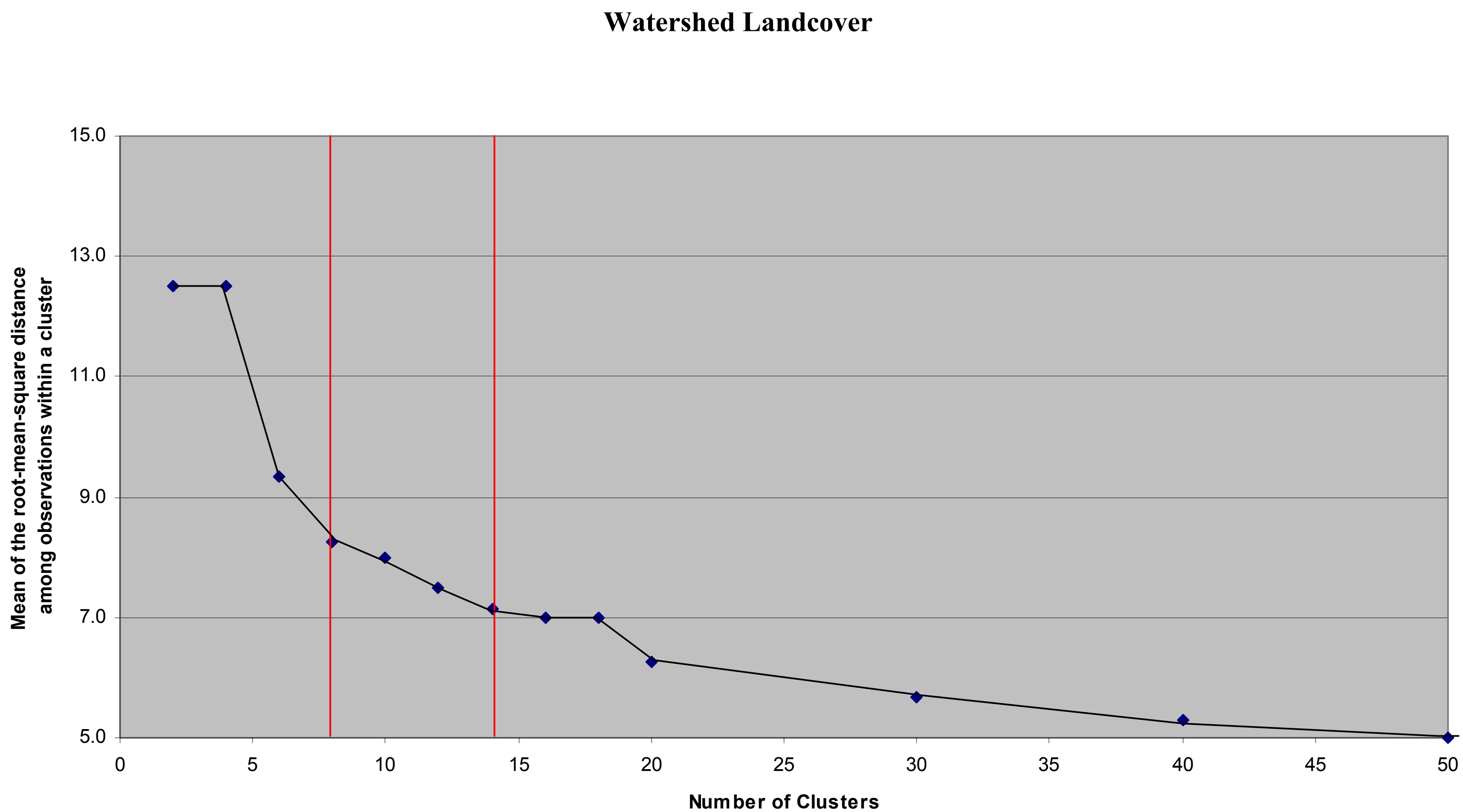


Figure 17. Line plot of the mean of the root-mean-square distance among observations within all clusters versus the number of clusters. This plot suggests that somewhere between eight and fourteen clusters only minimal additional variation, in the variables used to form the clusters (watershed land cover), is accounted for.

Figure 18. Line plot of the mean distance between cluster centroids versus the number of clusters. This plot suggests that above sixteen clusters only minimal additional variation, in the variables used to form the clusters (watershed land cover), is accounted for.

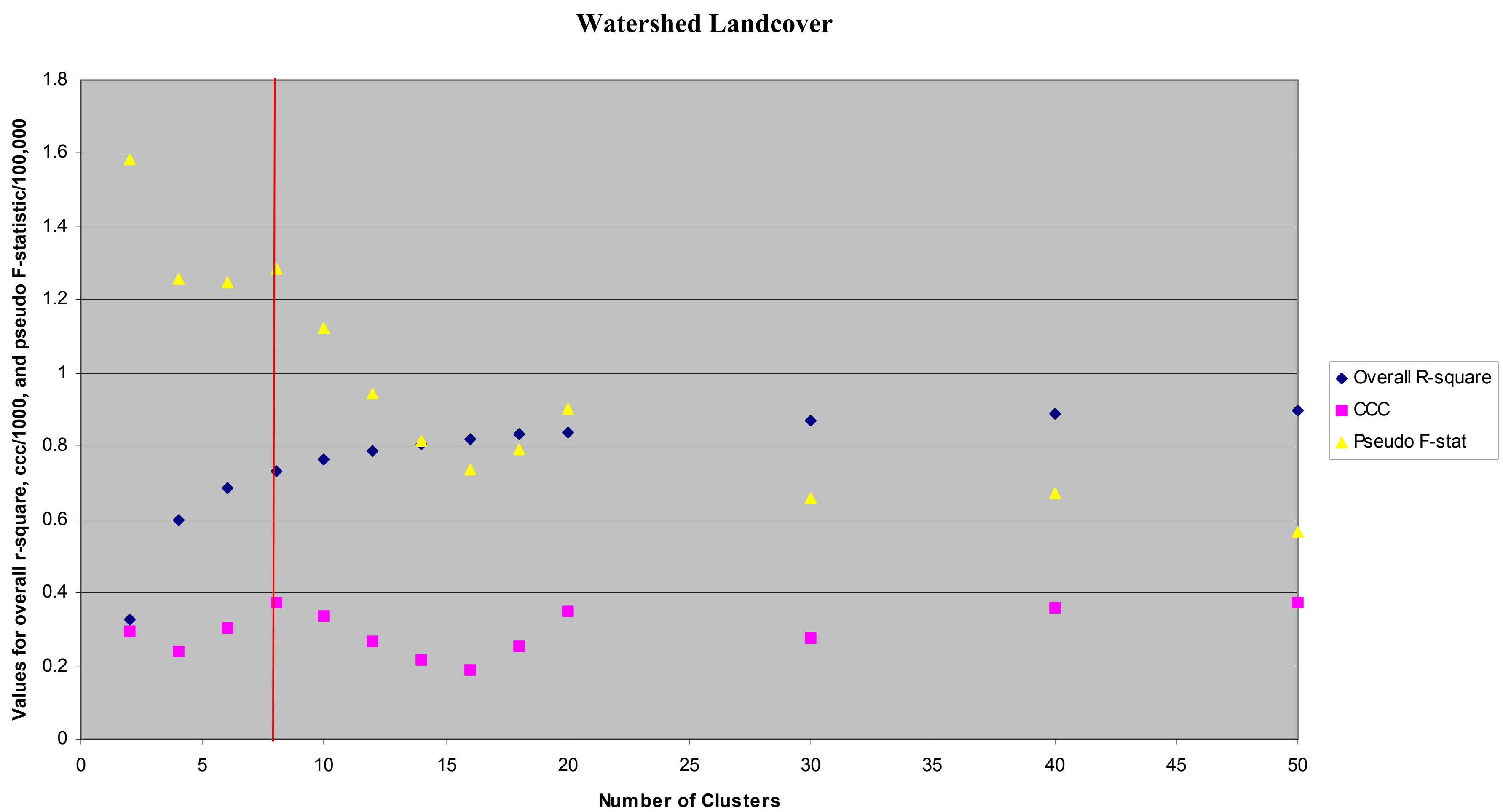


Figure 19. Plot of the overall r-square, cubic clustering criterion (CCC), and pseudo F-statistic values versus the number of clusters. This plot suggests that approximately 8 clusters is optimal for identifying relatively homogenous groupings for the six-class land cover. Note: for presentation purposes, the CCC and Psuedo F-statistic values were divided by 10,000 and 100,000, respectively.

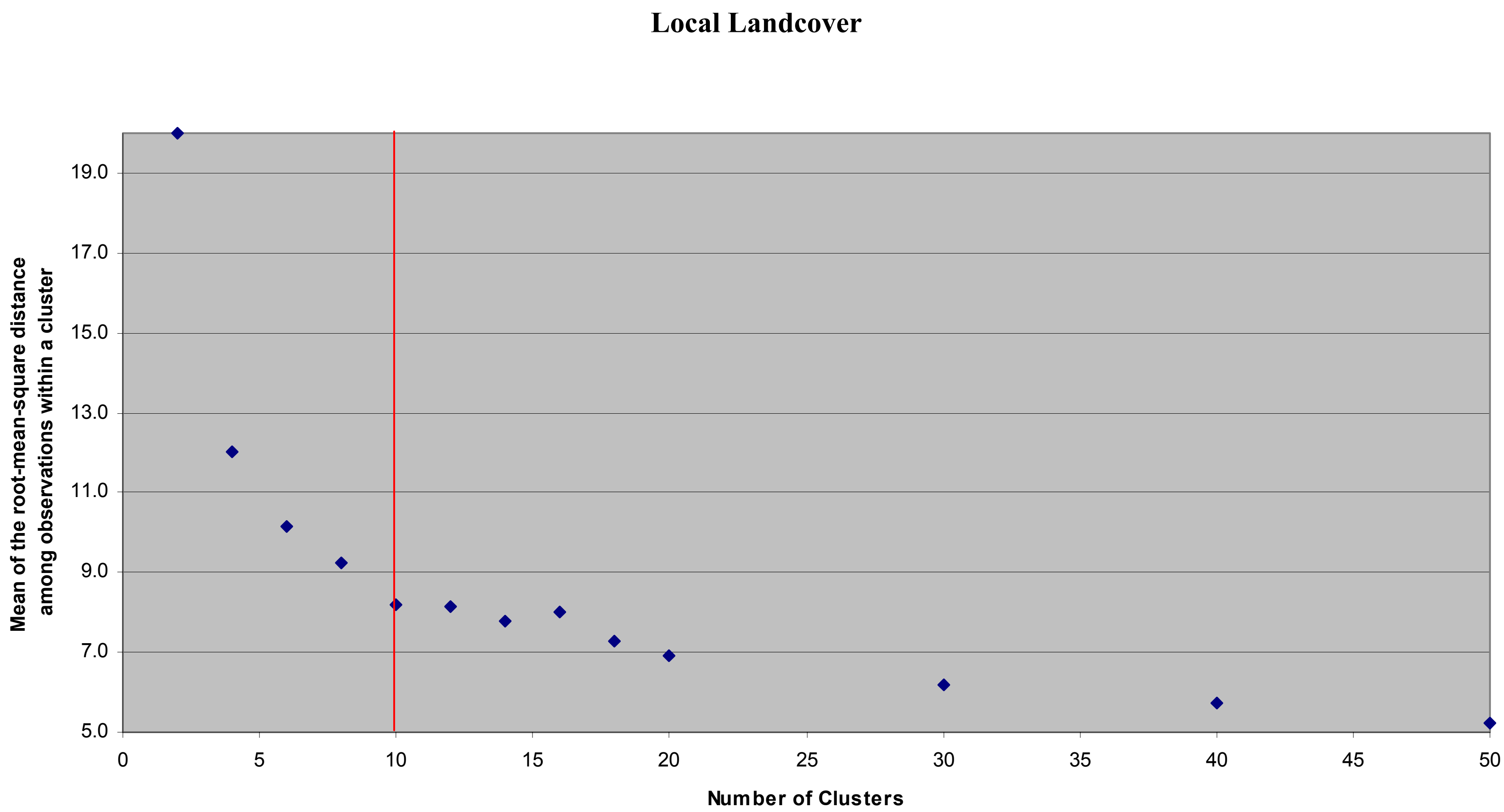


Figure 20. Line plot of the mean of the root-mean-square distance among observations within all clusters versus the number of clusters. This plot suggests that beyond ten clusters only minimal additional variation, in the variables used to form the clusters (local land cover), is accounted for.

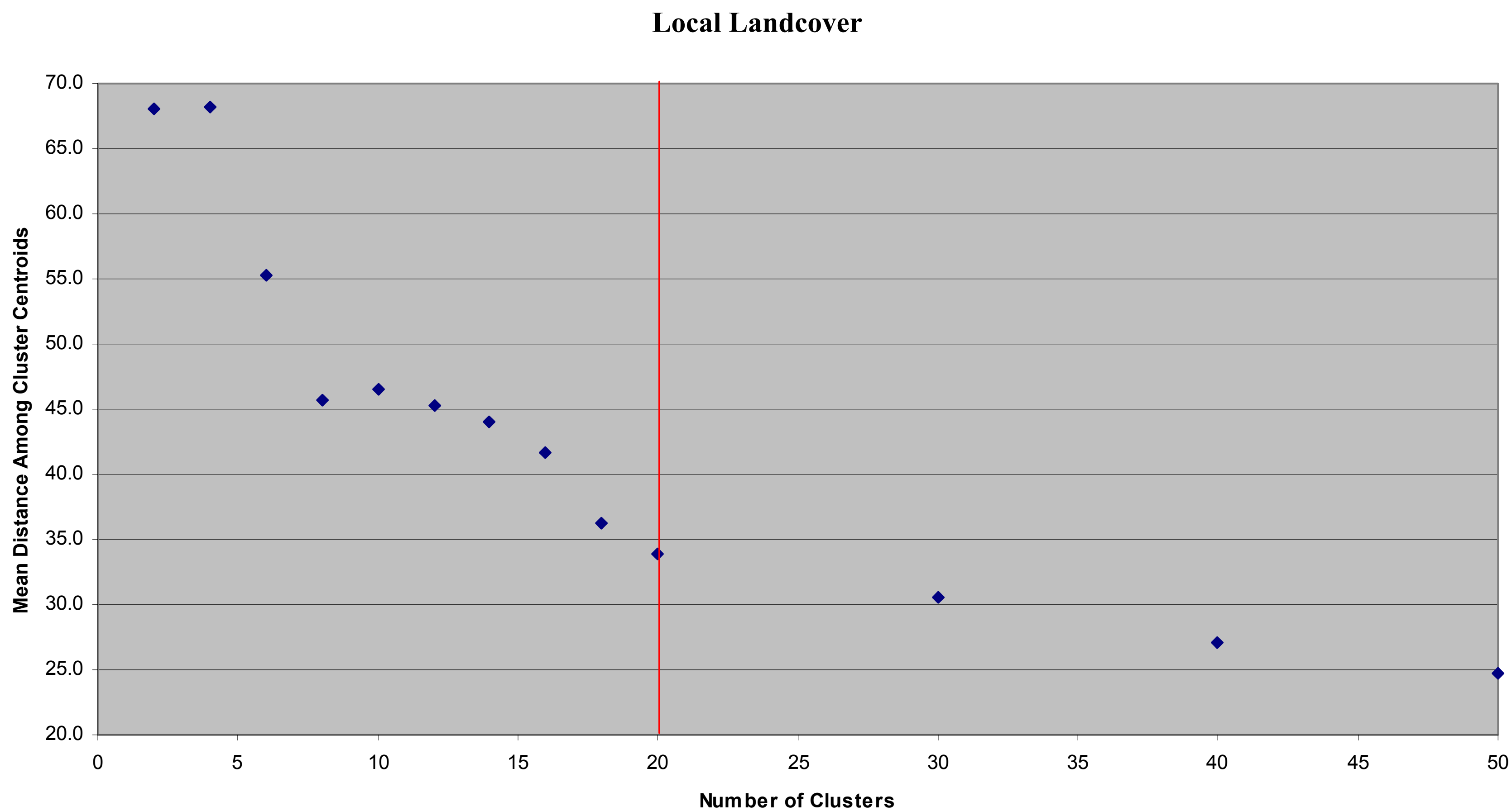


Figure 21. Line plot of the mean distance between cluster centroids versus the number of clusters. This plot suggests that above twenty clusters only minimal additional variation, in the variables used to form the clusters (watershed land cover), is accounted for.

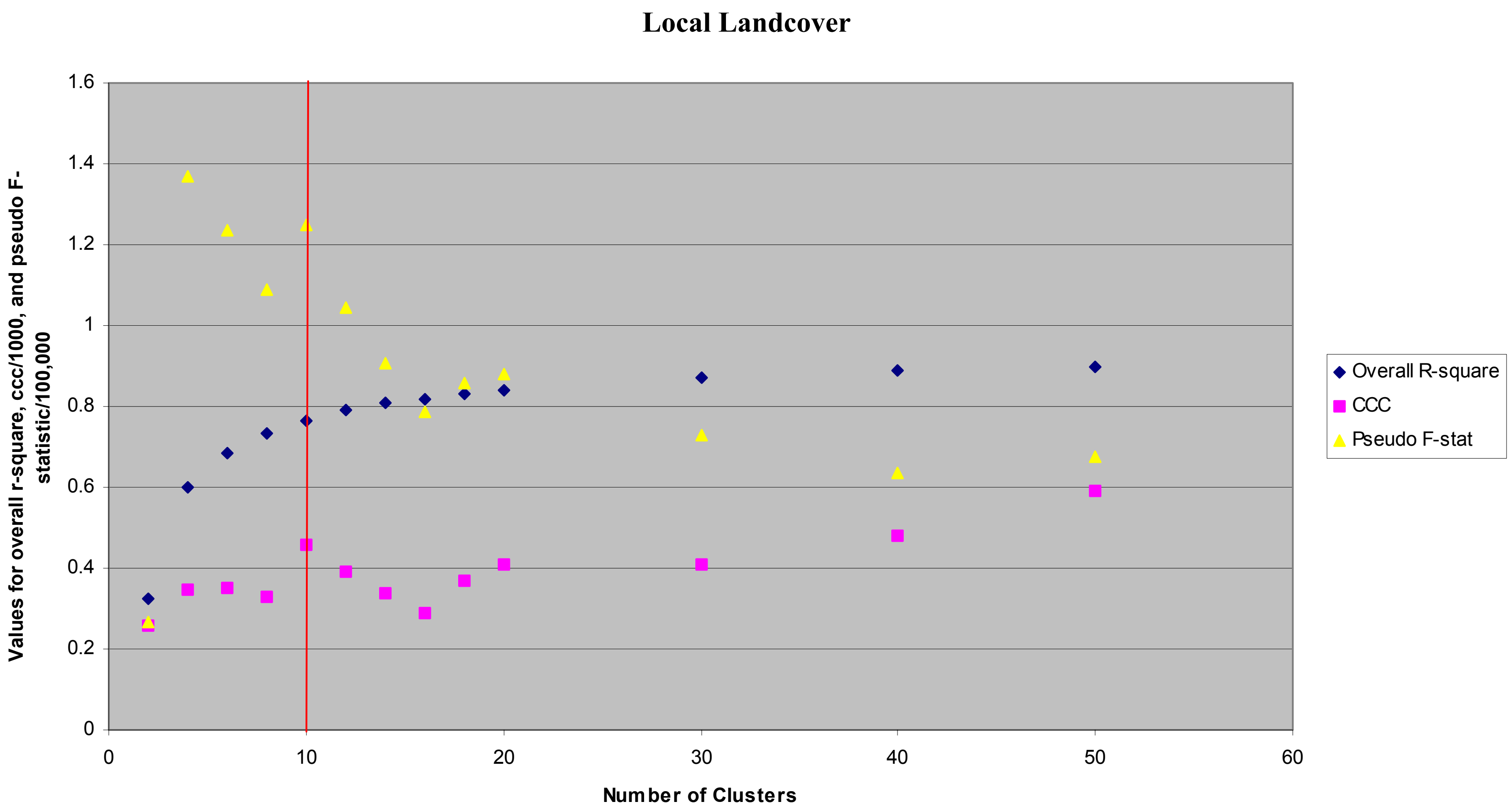


Figure 22. Plot of the overall r-square, cubic clustering criterion (CCC), and pseudo F-statistic values versus the number of clusters. This plot suggests that approximately 10 clusters is optimal for identifying relatively homogenous groupings for the six-class land cover (local). Note: for presentation purposes, the CCC and Psuedo F-statistic values were divided by 10,000 and 100,000, respectively.

Task 6 and 7 Important Information for Designing Research Projects

In addition to the diagnostic statistics, discussed above, the FASTCLUS procedure in SAS provides four important pieces of information that end users must be familiar with in order to fully utilize the resulting output within a GIS. First, each observation (i.e., stream segment) in the dataset is given a cluster number that ranges from 1 to N, where N equals the number of clusters. For instance, when generating just two clusters, all of the records in the dataset are coded as either 1 or 2 however; with 14 clusters the records in the dataset are given a cluster number of 1 through 14. SAS also provides a frequency value, which represents the total number of observations within the cluster. The frequency value is important because, from a research design perspective, it gives you a general indication of the potential for replication. Clusters with a relatively high number of observations will generally offer the best opportunity for finding replicate treatment and control sites, especially in those instances where additional post stratifications will be applied to the initial pool of potential candidate sites (e.g., further stratification based on higher resolution geospatial datasets like, SSURGO soils).

Although informative, the cluster number and frequency values still do not provide sufficient information for devising future research projects. These values only tell you which records are relatively similar to one another within the dataset and how many observations that are within a given cluster, but they tell you nothing about what these values actually represent. To understand the “character” of a specific cluster number you must refer to the cluster means tables (Tables 5-16). A cluster means table provides the mean values, within each cluster, for each of the variables used to generate the clusters. For instance, Table 8 provides the cluster means tables for the12 clusters generated from the reduced set of variables. For cluster number 1, in the 12 cluster set, there are 18,029 stream segments and these segments generally have watersheds dominated by local relief in the range of 51-100 feet (71.8% of their area) and soils falling within Hydrologic Soil Group D (88.7% of their area). These cluster mean tables were only generated for the cluster sets that would most likely be selected for designing research projects for each respective dataset based on the above diagnostic plots. For instance, for the reduced set of variables we produced cluster means tables for the 12, 14, and 16 cluster sets since the diagnostic plots suggested that approximately 12 to 16 was optimal (See Figures 14-16).

With the large number of stream segments included in the cluster analyses for this project, most clusters are comprised of thousands of individual segments. There are different degrees of variation among observations within each cluster, but the important point is that they are not completely homogenous and in fact in some instances observations from two extremes of the cluster could be quite different. To help overcome this problem the FASTCLUS provides a distance measure of each observation from the cluster centroid, which can be used as a guide to help select those observations that most closely resemble the values depicted in the cluster means table. Since the original distance measures provided by the FASTCLUS procedure represent distances in multivariate space, they are quite awkward to deal with. However, by ranking these distances from 1 to n, where n equals the number of observations within

Table 5. Cluster means table for the 14 clusters produced with the Full Set of watershed landscape variables. Bold values show the dominant landscape features for each landscape factor within each cluster.

		System Geology					
Cluster	Frequency	Devonian	Mississippian	Ordovician	Pennsylvanian	Quaternary	Silurian
1	5350	0.0	1.7	0.1	**97.8**	0.5	0.0
2	5528	0.9	6.8	**88.0**	4.0	0.3	0.0
3	20718	0.4	**41.0**	2.5	**55.7**	0.3	0.0
4	16470	0.0	15.4	0.8	**83.7**	0.0	0.0
5	3206	0.0	3.5	2.0	1.4	**93.0**	0.1
6	18688	0.0	2.2	0.1	**97.2**	0.4	0.0
7	3153	6.5	**52.7**	**28.6**	9.8	0.2	2.2
8	17943	0.2	4.1	0.6	**95.0**	0.0	0.1
9	2443	0.0	3.3	0.8	7.9	**87.7**	0.0
10	14490	0.8	**93.1**	1.4	4.1	0.4	0.1
11	13157	0.0	0.3	0.0	**99.1**	0.4	0.0
12	6199	1.8	**15.3**	**77.3**	2.4	1.9	1.3
13	1928	7.5	**22.0**	**57.7**	6.2	3.7	2.9
14	41072	0.0	0.3	0.0	**99.6**	0.1	0.0

		Relief					
Cluster	Frequency	0-50ft	51-100ft	101-200ft	201-300ft	301-500ft	501-700ft
1	5350	**67.4**	**30.9**	1.3	0.3	0.0	0.0
2	5528	10.2	**56.7**	**30.1**	2.9	0.1	0.0
3	20718	11.2	**33.5**	**54.6**	0.6	0.0	0.0
4	16470	**92.2**	7.5	0.3	0.0	0.0	0.0
5	3206	**90.3**	2.5	5.6	1.3	0.3	0.0
6	18688	10.4	**66.5**	**22.1**	0.9	0.1	0.0
7	3153	1.2	12.0	**73.1**	12.6	1.2	0.0
8	17943	**23.5**	**62.4**	14.1	0.1	0.0	0.0
9	2443	**82.7**	3.6	10.4	2.6	0.8	0.0
10	14490	13.5	**54.9**	**30.1**	1.4	0.2	0.0
11	13157	10.6	**35.4**	**53.8**	0.3	0.0	0.0
12	6199	1.9	14.8	**77.8**	4.5	1.0	0.0
13	1928	1.5	1.6	**18.8**	**71.3**	6.7	0.0
14	41072	5.3	**71.8**	**22.9**	0.0	0.0	0.0

Table 5. Continued.

Cluster	Frequency	Surface Texture Clayey	Cherty	Stony	Sandy	Loamy	Hydrologic Soil Group HSG_B	HSG_C	HSG_D
1	5350	2.7	0.1	0.6	0.0	**96.4**	0.8	**97.3**	1.6
2	5528	0.1	1.5	0.1	0.0	**98.2**	8.3	**85.3**	6.3
3	20718	1.0	0.2	0.5	0.1	**98.1**	3.9	**93.7**	1.7
4	16470	0.3	0.0	0.2	0.1	**99.2**	2.1	1.2	**96.6**
5	3206	**90.0**	0.0	0.1	0.3	5.8	**39.7**	14.4	**42.0**
6	18688	1.7	0.5	0.5	6.6	**90.5**	**92.7**	1.9	5.1
7	3153	0.1	0.2	**88.0**	1.2	10.2	**92.0**	1.9	5.9
8	17943	3.3	0.0	1.1	0.9	**94.5**	6.7	7.3	**84.8**
9	2443	3.3	0.0	0.0	0.7	**93.3**	**91.6**	4.7	0.9
10	14490	0.9	2.0	5.5	0.3	**90.8**	**53.1**	7.0	**39.0**
11	13157	**78.0**	0.3	0.1	3.1	**18.4**	**92.5**	6.7	0.6
12	6199	1.9	**28.7**	2.3	0.1	**65.7**	**89.7**	5.5	3.4
13	1928	2.2	0.9	**31.6**	0.0	**64.4**	**92.8**	4.1	2.3
14	41072	**38.2**	0.0	0.2	0.0	**61.5**	2.4	**96.6**	0.9

Table 6. Cluster means table for the 16 clusters produced with the Full Set of watershed landscape variables. Bold values show the dominant landscape features for each landscape factor within each cluster.

		System Geology					
Cluster	Frequency	Devonian	Mississippian	Ordovician	Pennsylvanian	Quaternary	Silurian
1	5435	0.6	3.0	**93.7**	2.7	0.0	0.0
2	12915	0.3	4.5	5.9	**88.6**	0.5	0.1
3	18007	0.2	2.7	0.1	**97.0**	0.0	0.0
4	10174	0.1	1.1	0.2	**97.6**	0.9	0.0
5	13997	0.2	3.1	1.9	**94.6**	0.1	0.0
6	338	0.1	**25.8**	0.6	**46.1**	23.3	0.0
7	5022	0.0	1.2	1.3	1.8	**95.6**	0.0
8	16828	0.0	0.0	0.0	**99.9**	0.1	0.0
9	1220	1.0	12.0	**41.9**	**39.6**	4.4	0.9
10	20975	0.2	**34.7**	0.9	**64.1**	0.0	0.1
11	10217	0.0	0.0	0.0	**99.9**	0.0	0.0
12	3407	0.0	**25.9**	**71.3**	2.5	0.4	0.0
13	10794	1.5	**83.5**	7.1	5.1	2.1	0.6
14	35180	0.1	**23.7**	0.2	**75.9**	0.1	0.0
15	2064	4.9	**13.6**	**50.0**	**17.6**	**10.6**	3.3
16	3772	6.8	**43.8**	**40.5**	6.8	0.2	1.8

		Relief					
Cluster	Frequency	0-50ft	51-100ft	101-200ft	201-300ft	301-500ft	501-700ft
1	5435	9.6	**48.5**	**38.0**	3.8	0.2	0.0
2	12915	**10.4**	**79.4**	9.9	0.2	0.0	0.0
3	18007	**25.5**	**61.5**	12.9	0.1	0.0	0.0
4	10174	7.2	**29.7**	**62.0**	1.0	0.0	0.0
5	13997	1.1	**17.9**	**80.3**	0.7	0.0	0.0
6	338	**24.6**	**60.5**	14.0	0.4	0.2	0.0
7	5022	**91.8**	1.8	5.0	1.1	0.3	0.0
8	16828	6.3	**55.1**	**38.5**	0.0	0.0	0.0
9	1220	**22.1**	**50.3**	**26.3**	1.2	0.2	0.0
10	20975	**76.1**	**21.1**	2.8	0.0	0.0	0.0
11	10217	13.1	**48.4**	**38.4**	0.1	0.0	0.0
12	3407	1.1	12.7	**84.1**	2.1	0.0	0.0
13	10794	7.9	**38.2**	**50.9**	2.7	0.2	0.0
14	35180	**19.0**	**74.0**	7.0	0.0	0.0	0.0
15	2064	6.8	2.8	**29.9**	**52.8**	7.7	0.0
16	3772	1.2	12.6	**60.0**	**23.9**	2.4	0.0

Table 6. Continued.

Cluster	Frequency	Surface Texture: Clayey	Cherty	Stony	Sandy	Loamy	Hydrologic Soil Group: HSG_B	HSG_C	HSG_D
1	5435	0.0	2.5	0.3	0.0	**97.1**	12.7	**85.6**	1.7
2	12915	0.8	0.2	0.2	0.6	**98.1**	**92.6**	3.2	4.0
3	18007	1.4	0.0	1.0	1.2	**96.2**	8.2	6.6	**84.1**
4	10174	10.9	0.7	1.8	**15.1**	**71.3**	**90.8**	4.2	4.3
5	13997	3.9	0.0	0.2	0.0	**95.8**	2.4	**96.3**	0.7
6	338	**32.4**	2.6	2.5	1.2	**47.8**	**20.4**	**47.2**	**18.9**
7	5022	**56.5**	0.0	0.1	0.5	**39.4**	**61.1**	11.0	**24.3**
8	16828	**84.2**	0.0	0.3	0.0	15.4	2.2	**96.7**	0.9
9	1220	**50.6**	0.0	1.2	0.0	**47.8**	4.3	0.2	**95.0**
10	20975	0.1	0.0	0.9	0.1	**98.8**	2.8	2.3	**94.8**
11	10217	**96.8**	0.0	0.0	0.0	3.2	**89.5**	10.4	0.0
12	3407	0.9	**55.8**	0.4	0.1	**40.3**	**94.7**	2.8	0.1
13	10794	1.1	1.6	6.4	0.7	**89.6**	**80.0**	7.8	11.1
14	35180	1.9	0.1	0.1	0.0	**97.8**	1.3	**96.9**	1.7
15	2064	3.1	0.0	2.7	0.0	**93.3**	**94.9**	2.6	1.7
16	3772	0.2	0.0	**89.0**	0.0	10.7	**93.8**	1.5	4.6

Table 7. Cluster means table for the 18 clusters produced with the Full Set of watershed landscape variables. Bold values show the dominant landscape features for each landscape factor within each cluster.

		System Geology					
Cluster	Frequency	Devonian	Mississippian	Ordovician	Pennsylvanian	Quaternary	Silurian
1	2730	0.1	1.6	1.1	**18.9**	**78.1**	0.0
2	11862	0.1	0.5	2.1	**97.2**	0.0	0.0
3	16535	0.0	0.0	0.0	**99.9**	0.1	0.0
4	7356	1.0	**91.7**	1.9	3.9	1.1	0.5
5	15282	0.2	**27.5**	0.8	**71.5**	0.0	0.0
6	1573	0.0	**19.6**	**76.0**	4.4	0.0	0.0
7	8168	2.3	4.0	**48.2**	**40.9**	3.4	1.3
8	3196	0.0	4.0	2.0	1.6	**92.2**	0.1
9	3722	4.2	**78.2**	6.9	8.9	0.3	1.5
10	30950	0.0	1.4	0.1	**98.4**	0.1	0.0
11	9855	0.0	0.0	0.0	**99.7**	0.1	0.0
12	1626	7.3	6.4	**81.3**	3.4	0.6	1.0
13	6187	0.4	**95.6**	0.5	3.2	0.2	0.0
14	5074	0.8	2.7	**93.0**	3.0	0.6	0.0
15	18863	0.1	1.5	0.1	**98.3**	0.0	0.0
16	17927	0.0	2.2	0.2	**97.3**	0.3	0.0
17	7491	0.5	**93.7**	1.0	4.4	0.4	0.0
18	1948	3.0	**27.3**	**28.1**	**40.3**	0.0	1.1

		Relief					
Cluster	Frequency	0-50ft	51-100ft	101-200ft	201-300ft	301-500ft	501-700ft
1	2730	**86.0**	3.6	6.8	2.9	0.7	0.0
2	11862	0.9	12.7	**85.9**	0.6	0.0	0.0
3	16535	6.2	**55.3**	**38.4**	0.0	0.0	0.0
4	7356	8.6	**40.7**	**43.8**	6.4	0.5	0.0
5	15282	**88.3**	10.0	1.7	0.0	0.0	0.0
6	1573	1.9	**23.1**	**72.4**	2.7	0.0	0.0
7	8168	2.2	**10.4**	**77.4**	8.5	1.5	0.0
8	3196	**90.0**	2.7	5.5	1.4	0.3	0.0
9	3722	1.8	**18.0**	**73.1**	6.8	0.3	0.0
10	30950	**15.5**	**74.4**	10.0	0.1	0.0	0.0
11	9855	12.9	**44.6**	**42.4**	0.1	0.0	0.0
12	1626	1.4	6.4	**40.4**	**46.5**	5.3	0.0
13	6187	9.9	**79.6**	**10.5**	0.0	0.0	0.0
14	5074	10.8	**50.5**	**34.6**	3.9	0.2	0.0
15	18863	**34.3**	**53.3**	12.4	0.1	0.0	0.0
16	17927	9.0	**70.2**	**19.8**	1.0	0.1	0.0
17	7491	**29.3**	**51.0**	**19.3**	0.4	0.0	0.0
18	1948	7.7	**76.1**	**15.8**	0.3	0.0	0.0

Table 7. Continued

Cluster	Frequency	Clayey	Cherty	Stony	Sandy	Loamy	HSG_B	HSG_C	HSG_D
		Surface Texture					Hydrologic Soil Group		
1	2730	4.0	0.0	0.1	0.6	**92.8**	**91.4**	4.7	1.3
2	11862	6.9	0.0	0.2	0.0	**92.7**	3.2	**95.1**	0.8
3	16535	**89.3**	0.0	0.2	0.0	10.4	4.3	**93.1**	2.5
4	7356	1.0	0.9	0.2	0.2	**97.3**	**92.2**	3.3	3.5
5	15282	0.1	0.0	0.7	0.0	**99.2**	1.6	1.4	**97.0**
6	1573	0.4	**87.1**	0.0	0.0	8.9	**89.1**	7.3	0.1
7	8168	2.3	0.4	2.4	0.2	**94.5**	**91.4**	5.1	3.0
8	3196	**90.5**	0.0	0.1	0.3	5.3	**39.0**	14.0	**43.2**
9	3722	0.7	**15.8**	**63.2**	2.1	**16.9**	**84.5**	2.1	12.1
10	30950	4.1	0.0	0.2	0.0	**95.5**	1.7	**96.1**	1.7
11	9855	**95.3**	0.0	0.0	0.0	4.6	**95.9**	4.0	0.0
12	1626	0.7	0.0	**86.6**	0.0	12.5	**96.2**	1.6	2.1
13	6187	0.7	1.3	4.5	0.1	**93.0**	13.3	**32.9**	**52.9**
14	5074	0.1	2.0	0.2	0.0	**97.6**	9.5	**89.2**	1.3
15	18863	2.2	0.0	1.0	0.9	**95.8**	6.9	5.5	**87.1**
16	17927	2.7	0.5	0.7	9.1	**86.9**	**91.6**	2.4	5.5
17	7491	0.3	0.3	0.2	0.0	**99.1**	2.6	**96.1**	1.3
18	1948	0.5	0.1	2.5	0.4	**96.3**	8.2	6.1	**85.3**

Table 8. Cluster means table for the 12 clusters produced with the Reduced Set of watershed landscape variables. Bold values show the dominant landscape features for each landscape factor within each cluster.

		Relief						Hydrologic Soil Group		
Cluster	**Frequency**	**0-50ft**	**51-100ft**	**101-200ft**	**201-300ft**	**301-500ft**	**501-700ft**	**HSG_B**	**HSG_C**	**HSG_D**
1	18029	**17.8**	**71.8**	10.3	0.0	0.0	0.00	5.6	5.6	**88.7**
2	5883	**86.2**	**9.7**	3.3	0.6	0.1	0.00	**93.0**	1.7	3.4
3	2166	5.5	**16.0**	**73.8**	4.4	0.3	0.00	17.3	1.9	**80.5**
4	20549	**87.9**	**11.2**	0.8	0.0	0.0	0.00	2.4	1.7	**95.6**
5	310	2.4	1.0	9.9	**28.9**	**57.8**	0.00	**96.4**	1.4	1.2
6	371	0.4	0.5	**27.5**	**70.1**	1.4	0.00	14.4	**85.6**	0.0
7	7736	**75.4**	**23.0**	1.6	0.0	0.0	0.00	2.6	**92.8**	3.9
8	21981	1.9	**15.4**	**80.3**	2.3	0.1	0.00	**93.9**	2.7	2.6
9	43094	7.8	**81.9**	**10.3**	0.0	0.0	0.00	2.3	**96.0**	1.6
10	2702	2.8	1.9	**26.4**	**65.7**	3.2	0.00	**96.5**	1.6	1.1
11	24418	8.1	**78.4**	**13.4**	0.1	0.0	0.01	**88.2**	4.9	5.8
12	23106	1.2	**18.8**	**79.5**	0.4	0.0	0.00	4.9	**93.4**	1.0

Table 9. Cluster means table for the 14 clusters produced with the Reduced Set of watershed landscape variables. Bold values show the dominant landscape features for each landscape factor within each cluster.

		Relief						Hydrologic Soil Group		
Cluster	Frequency	0-50ft	51-100ft	101-200ft	201-300ft	301-500ft	501-700ft	HSG_B	HSG_C	HSG_D
1	5941	**85.9**	**10.0**	3.4	0.6	0.1	0.0	**93.2**	1.8	2.9
2	156	13.9	7.3	**31.7**	**44.6**	2.5	0.0	**41.4**	0.2	**56.8**
3	20418	**88.2**	**11.1**	0.7	0.0	0.0	0.0	2.1	1.8	**95.8**
4	3839	4.4	**25.4**	**69.7**	0.5	0.0	0.0	12.3	2.8	**82.3**
5	23408	1.1	**19.9**	**78.6**	0.4	0.0	0.0	4.1	**95.3**	0.5
6	7254	**76.7**	**22.2**	1.1	0.0	0.0	0.0	1.5	**97.1**	1.2
7	22249	1.9	**16.2**	**79.6**	2.2	0.1	0.0	**94.5**	2.9	1.8
8	359	0.4	0.5	**26.3**	**71.3**	1.5	0.0	15.0	**85.0**	0.0
9	2695	2.7	1.9	**26.4**	**65.8**	3.2	0.0	**96.5**	1.6	1.1
10	9526	**22.3**	**59.6**	17.7	0.4	0.0	0.0	**36.4**	**27.0**	**33.4**
11	19452	6.2	**82.7**	**11.0**	0.0	0.0	0.0	**95.9**	2.0	2.0
12	14027	**17.2**	**76.9**	5.9	0.0	0.0	0.0	3.0	3.1	**93.8**
13	40710	7.5	**83.2**	**9.3**	0.0	0.0	0.0	1.8	**97.0**	1.1
14	311	2.5	1.1	9.9	**28.8**	**57.8**	0.0	**96.3**	1.4	1.4

Table 10. Cluster means table for the 16 clusters produced with the Reduced Set of watershed landscape variables. Bold values show the dominant landscape features for each landscape factor within each cluster.

		Relief						Hydrologic Soil Group		
Cluster	Frequency	0-50ft	51-100ft	101-200ft	201-300ft	301-500ft	501-700ft	HSG_B	HSG_C	HSG_D
1	318	2.6	0.9	9.9	**29.8**	**56.8**	0.0	**97.6**	0.7	1.0
2	6761	12.9	**68.4**	**18.3**	0.3	0.0	0.0	**31.6**	**47.4**	17.5
3	5155	**38.5**	**46.9**	13.8	0.6	0.1	0.0	**42.6**	9.6	**46.1**
4	22502	1.0	**19.0**	**79.7**	0.4	0.0	0.0	3.4	**96.1**	0.4
5	90	8.8	4.7	**27.0**	**57.1**	2.3	0.0	**35.1**	0.4	**64.2**
6	744	11.5	**17.7**	**52.4**	**17.3**	1.1	0.0	**52.2**	**38.5**	3.3
7	21765	1.8	**15.8**	**80.1**	2.2	0.1	0.0	**95.1**	2.6	1.6
8	5388	**88.5**	**7.9**	2.9	0.5	0.1	0.0	**95.2**	1.4	1.5
9	308	0.2	0.4	**25.6**	**72.6**	1.3	0.0	9.1	**90.9**	0.0
10	3978	4.0	**25.4**	**70.1**	0.5	0.0	0.0	14.4	3.3	**79.2**
11	2597	2.6	1.9	**26.2**	**66.3**	3.1	0.0	**97.8**	0.7	1.0
12	19447	6.3	**82.4**	**11.3**	0.0	0.0	0.0	**96.2**	1.5	2.1
13	14401	**17.2**	**76.5**	6.3	0.0	0.0	0.0	3.2	3.3	**93.4**
14	19861	**88.9**	**10.5**	0.6	0.0	0.0	0.0	1.5	1.7	**96.5**
15	39740	7.4	**83.1**	**9.5**	0.0	0.0	0.0	1.1	**98.1**	0.8
16	7290	**76.6**	**22.3**	1.1	0.0	0.0	0.0	1.5	**97.0**	1.3

Table 11. Cluster means table for the 10 clusters produced with the Watershed Land Cover variables. Bold values show the dominant land cover classes for each cluster. Bold cluster numbers indicate those clusters best suited for research.

Cluster	Frequency	Urban	Cropland	Grassland	Forest	Water
1	586	0.8	11.3	**25.5**	**24.4**	**31.2**
2	31700	0.1	**81.3**	**14.0**	3.9	0.4
3	223	0.1	3.2	3.6	**15.3**	**74.7**
4	39813	0.4	**12.5**	**76.7**	10.1	0.4
5	9408	0.5	4.4	**18.0**	**75.6**	1.3
6	677	**72.0**	1.8	**14.9**	10.8	0.4
7	2252	**38.6**	7.2	**37.3**	16.3	0.5
8	7326	0.4	**42.8**	**22.1**	**32.9**	1.2
9	52946	0.4	**43.7**	**45.8**	9.5	0.4
10	25414	1.1	11.0	**48.3**	**38.7**	0.7

Table 12. Cluster means table for the 12 clusters produced with the Watershed Land Cover variables. Bold values show the dominant land cover classes for each cluster. Bold cluster numbers indicate those clusters best suited for research.

Cluster	Frequency	Urban	Cropland	Grassland	Forest	Water
1	39533	0.3	**12.5**	**76.7**	10.1	0.3
2	177	0.1	3.5	3.6	10.0	**79.8**
3	414	1.0	14.3	**37.1**	14.4	**26.8**
4	25431	0.9	11.2	**48.4**	**38.6**	0.7
5	1756	**35.7**	9.6	**43.7**	10.4	0.5
6	7810	0.4	**44.9**	**22.1**	**30.8**	1.1
7	960	**38.8**	1.9	**22.9**	**35.8**	0.6
8	31416	0.1	**81.4**	**14.0**	3.8	0.4
9	615	**73.7**	2.1	**16.2**	7.6	0.4
10	52441	0.4	**43.7**	**46.0**	9.4	0.4
11	9258	0.3	4.6	**18.2**	**75.9**	0.9
12	534	0.3	5.5	12.4	**45.9**	**30.0**

Table 13. Cluster means table for the 14 clusters produced with the Watershed Land Cover variables. Bold values show the dominant land cover classes for each cluster. Bold cluster numbers indicate those clusters best suited for research.

Cluster	Frequency	Urban	Cropland	Grassland	Forest	Water
1	310	1.0	10.9	**43.8**	15.2	**26.9**
2	499	0.3	3.9	13.0	**49.9**	**28.9**
3	6663	0.4	**39.6**	**23.0**	**35.8**	0.8
4	9118	0.3	4.3	**18.3**	**76.1**	0.9
5	32044	0.1	**81.1**	**14.4**	3.8	0.3
6	889	**35.5**	1.7	**23.1**	**39.2**	0.6
7	1753	**38.5**	8.9	**41.6**	10.5	0.5
8	598	**74.0**	1.9	**15.3**	8.3	0.4
9	224	1.2	**20.5**	17.6	**19.6**	**22.5**
10	24871	0.9	10.8	**49.1**	**38.4**	0.7
11	53278	0.4	**42.8**	**46.7**	9.6	0.4
12	177	0.1	2.8	3.7	10.7	**80.1**
13	38060	0.3	**12.1**	**77.2**	10.0	0.3
14	1861	1.0	**59.7**	**18.4**	14.5	4.8

Table 14. Cluster means table for the 10 clusters produced with the Local Land Cover variables. Bold values show the dominant land cover classes for each cluster. Bold cluster numbers indicate those clusters best suited for research.

Cluster	Frequency	Urban	Cropland	Grassland	Forest	Water
1	12497	0.2	**43.7**	18.6	**36.2**	0.8
2	30257	0.6	9.6	**47.3**	**41.8**	0.6
3	31212	0.1	**83.7**	11.2	4.3	0.5
4	2041	0.6	10.2	13.1	**26.8**	**37.1**
5	1967	**39.9**	5.8	**37.6**	15.9	0.7
6	39547	0.3	**44.3**	**45.7**	9.1	0.4
7	36695	0.3	10.0	**79.5**	9.9	0.3
8	711	**75.4**	1.6	12.9	9.3	0.7
9	1020	0.3	2.3	3.5	8.3	**84.2**
10	14398	0.3	4.0	13.8	**80.3**	1.2

Table 15. Cluster means table for the 12 clusters produced with the Local Land Cover variables. Bold values show the dominant land cover classes for each cluster. Bold cluster numbers indicate those clusters best suited for research.

Cluster	Frequency	Urban	Cropland	Grassland	Forest	Water
1	12403	0.17	**43.7**	18.7	**36.4**	0.6
2	1025	0.38	2.1	4.3	7.7	**83.7**
3	1980	**39.74**	5.8	**37.6**	16.0	0.9
4	714	**75.35**	1.6	13.0	9.3	0.7
5	31097	0.08	**83.7**	11.2	4.3	0.4
6	1509	0.47	4.2	12.9	**39.7**	**39.3**
7	724	0.48	**40.5**	10.9	10.1	**33.9**
8	39525	0.26	**44.3**	**45.8**	9.1	0.4
9	36656	0.26	9.9	**79.5**	9.9	0.3
10	14247	0.35	4.0	13.9	**80.5**	0.9
11	254	0.11	5.4	5.0	6.7	4.9
12	30211	0.56	9.6	**47.3**	**41.7**	0.6

Table 16. Cluster means table for the 10 clusters produced with the Local Land Cover variables. Bold values show the dominant land cover classes for each cluster. Bold cluster numbers indicate those clusters best suited for research.

Cluster	Frequency	Urban	Cropland	Grassland	Forest	Water
1	667	0.5	**44.3**	8.2	10.3	**32.2**
2	36622	0.2	10.0	**79.6**	9.9	0.3
3	689	1.6	8.7	**38.9**	14.9	**33.3**
4	1531	**39.2**	7.0	**43.0**	10.2	0.5
5	310009	0.1	**83.8**	11.2	4.3	0.4
6	30003	0.4	9.7	**47.4**	**41.8**	0.5
7	1307	0.3	4.2	9.4	**43.6**	**38.9**
8	14105	0.2	4.1	13.8	**80.7**	0.9
9	988	0.3	2.1	3.3	7.8	**84.7**
10	814	**37.5**	2.1	**20.4**	**39.3**	0.7
11	12364	0.1	**43.8**	18.6	**36.3**	0.6
12	239	0.0	5.0	3.5	6.4	3.9
13	635	**77.5**	1.7	13.6	6.5	0.7
14	39372	0.3	**44.4**	**45.8**	9.1	0.3

the cluster, the rankings become relativized values that can then be more easily compared among observations. Consequently, for each cluster set that was produced for each dataset, we ranked the distance values within each individual cluster from 1 to n and these values are provided in the associated GIS coverages discussed below. These cluster distance ranks indicate how far a given segment is from the cluster centroid. The smaller the value of the distance rank, the closer the segment is to the cluster centroid (i.e., mean). That is, the smaller the distance rank value the more closely the segment represents the landscape characteristics of the cluster that are presented in the cluster means tables.

The user must, however, be conscious of the fact that these ranks are relative to the number of observations within the cluster. A stream segment with distance rank of 1,000 within a cluster containing 40,000 observations might be very similar to values represented in the cluster means table, while a segment with a distance rank of 100 might be very different from the mean values if there are only 300 observations within the cluster. These distance ranks should therefore only be used as an initial guide to assessing the relative similarity of two stream segments. The only way to truly assess the similarity of two or more segments is to refer back to the raw input data that was used to generate the clusters, which is also discussed in the following section.

Task 8: Develop a GIS database that could be used to help develop experimental designs and select potential study sites for examining potential effects of headwater impoundments on the ecological integrity of Missouri streams

GIS Database Descriptions

All of the data developed for this project is contained within 10 separate ArcInfo coverages. These coverages represent the entire stream network within the study area and can be joined together using the ARCIDNUM item. There was a significant amount of data gathered and summarized for this project that was not ultimately used in the cluster analyses. However, since these data could assist with a wide variety of research and management efforts we did not want to simply discard this potentially valuable information. Consequently, we developed two coverages that include all of the raw data, as well as the data used for the cluster analyses.

Specifically;

- NETW_ALL
    - Contains the raw data pertaining to the overall watershed of each individual stream segment. Appendix A provides a detailed description of the attributes in this coverage.
- NETL_ALL
    - Contains the raw data pertaining to the individual segmentshed of each stream segment. Appendix B provides a detailed description of the attributes within this coverage.

There are four coverages that include the raw data for just those variables used in the cluster analyses. Specifically;

- CLUSTALL_VARS
  - Includes the data used to run the cluster analyses for the Full Set of watershed landscape variables (see Appendix C for attribute descriptions).
- CLUSTRED_VARS
  - Includes the data used to run the cluster analyses for the Reduced Set of watershed landscape variables (see Appendix D for attribute descriptions).
- CLUST_LULCW_VARS
  - Includes the data used to run the cluster analyses for the Watershed Land Cover Set of variables (see Appendix E for attribute descriptions).
- CLUST_LULCL_VARS
  - Includes the data used to run the cluster analyses for the Local Land Cover Set of variables (see Appendix F for attribute descriptions).

The remaining four coverages provide the results of the cluster analyses. Specifically;

- CLUST_ALL
  - Contains the cluster numbers and associated distance ranks for each observation, within each cluster set, generated for the Full Set of watershed variables (see Appendix G for attribute descriptions).
- CLUST_RED
  - Contains the cluster numbers and associated distance ranks for each observation, within each cluster set, generated for the Reduced Set of watershed variables (see Appendix H for attribute descriptions).
- CLUST_LULCW
  - Contains the cluster numbers and associated distance ranks for each observation, within each cluster set, generated for the Watershed Land Cover Set (see Appendix I for attribute descriptions).
- CLUST_LULCL
  - Contains the cluster numbers and associated distance ranks for each observation, within each cluster set, generated for the Local Land Cover Set (see Appendix J for attribute descriptions).

Using the GIS databases

In order to use the resulting GIS databases for assessing experimental design options or selecting specific study sites, the user must first answer a series of six questions. Each of these questions is listed below in chronological order and when possible, recommended options are provided based upon what we have learned through constructing and working with these databases.

1. Which cluster datasets are you going to use?

In attempting to account for inherent natural variation among potential study sites, the user can either elect to use the results from the cluster analyses for the Full (CLUST_ALL) or Reduced (CLUST_RED) set of landscape variables. Because the full set contains more variables and landscape factors (e.g., soil surface texture and geology) than the Reduced Set, it should, in theory, account for a higher degree of the natural variation between potential study sites than the Reduced Set. However, considering that much of northern Missouri is covered with a mantle of glacial till, the variation that is accounted for with the inclusion of System-level geology in the Full Set may not account for any meaningful inherent natural differences among streams within the study area. Yet, many of the headwater streams and some of the larger streams in north Missouri have cut entirely through this mantle of glacial till (VanDike 1979; Pflieger 1997; Nigh and Schroder 2002) and as consequence the geomorphic and instream habitat properties of these stream channels are to some degree influenced by bedrock geology. Also, studies within the Lower Peninsula of Michigan, which is also covered by a thick layer of glacial till, have found that geographic variations in stream water chemistry and habitat are associated with geographic variations in bedrock geology (Richards et al. 1996; Wang et al. 2003).

Again, because the Full Set contains more variables than the Reduced Set the diagnostic plots indicate that more clusters must be used with the Full Set in order to fully account for variation among stream segments. Not surprising, as you increase the number of clusters the number of observations within each cluster decreases, which diminishes replication potential. Consequently, we have found that the replication potential of the Reduced Set is marginally better for headwater streams, but becomes substantially better as stream size increases.

Ultimately, the decision to use the cluster results from the Full or Reduced Set of variables will have to be determined by all interested parties. The Full Set potentially accounts for more inherent natural variation among possible study sites, but using this set may reduce replication potential, since there are generally fewer observations within each cluster of the Full Set.

To account for land use/cover differences among potential study sites the user can select either the cluster results for the Watershed Land Cover or Local Land Cover, or both. We recommend using both since watershed and local land cover/use influences the biological integrity of streams (Lammert and Allan 1999, Wang et al. 2003). By using both the user saves time in selecting potential study sites that are comparable with regard to general watershed and local land cover characteristics.

2. How many clusters are you going to use in the selected datasets?

The diagnostic plots created in Tasks 6 and 7 (See Figures 11-22) should be used to guide this decision. Again, this decision will have to be made by all interested parties. The only guidance we can provide is that as you increase the number of clusters, the number of observations within each cluster decreases, which diminishes replication potential. However, once question 1 is answered it might be best to examine replication options with multiple numbers of clusters. For example, assess the replication potential with the results from the 14, 16, and 18 clusters generated for the Watershed Full Set and then use the highest number of clusters that still provides sufficient replication options.

3. What specific clusters are you going to focus on within the selected cluster set for each of the datasets?

Not all individual clusters are well suited to examining the potential effects of headwater impoundments on the biological integrity of northern Missouri streams. For instance, some clusters contain relatively few observations because they represent rare landscape or rare land cover conditions within the study area. Consequently, the user should examine frequency values within the appropriate cluster means tables (See Tables 5-16) as an initial guide for identifying clusters that contain a relatively high number of stream segments. This selection criterion simply relates to the fact that with a higher number of segments there will be a higher probability of finding replicates with and without headwater impoundments.

Another thing to consider when answering this question is the relative homogeneity of the cluster, which can also be ascertained from the cluster means tables. For instance, if possible, clusters that are dominated by one or two geologic, soil, relief, or land cover classes should be selected in order to minimize variation in landscape character among potential study sites, which may confound research findings.

The user will also want to use the appropriate cluster means tables to further identify those clusters that are not dominated by an undesired watershed or local land cover type (e.g., water or urban). Those with a high percentage of water are mainly inundated by larger reservoirs and are unsuited for study in relation to the overall goal of the broader project. Since the effects of urban land can be so pervasive and unpredictable, segments with a high percentage of this land cover class, locally or within their watershed, would also not likely be suited for assessing the potential effects of headwater impoundments.

4. What range of stream size/drainage area are you going to focus on?

First of all, we recommend using a range of Shreve link values for designating stream size classes because it is a more accurate measure of stream size than Strahler order and because it is easier to work with than drainage area. Specifically, for headwaters, where the desired experimental design is to examine potential individual effects of headwater impoundments (treatment vs. control), we recommend a link value range 1 to 10. We recommend this range because our queries have shown that headwater impoundments almost always (over 95%) occur on streams falling within this range of values.

For larger streams, where the desired experimental design is to examine possible cumulative effects of multiple headwater impoundments, the end user must specify their desired range of values. The only guidance we can provide in this instance is to try, to the extent possible, to minimize differences among the link values of selected sites (e.g., within 50 to 200) in order to minimize inherent differences resulting from natural longitudinal variation, which may obscure research findings.

5. What size criteria are you going to use to define headwater impoundments?

We recommend using the 2-15 acre size criteria since we have found it more accurately identifies headwater impoundments than the 2-50 acre size criteria, which tends to identify too many larger impoundments.

6. How are you going to identify the presence of a headwater impoundment?

We recommend using the 100-meter tolerance since it more accurately captures headwater impoundments than the no tolerance option. However, even when using the tolerance some headwater impoundments are missed so subsequent visual examination of the data against the backdrop of the water body coverages will always be necessary during the more detailed site-selection process.

Once these questions are answered, the user can follow the specific steps listed below in order to assess experimental design options and begin the process of selecting an initial set of potential study sites.

1. Open the selected cluster dataset (e.g., CLUST_ALL) in ArcGIS.

2. Using the chosen number of clusters for the landscape dataset (e.g., 14 clusters from CLUST_ALL) and the chosen number of clusters for the watershed and/or local land cover datasets (e.g., 10 clusters for each), concatenate the cluster numbers across those specific data columns.
   - First, add a new item to the table (e.g., ALL14_CAT) to hold the concatenated values.

   - Open the Field Calculator and use the appropriate expression (similar to the one that follows) to calculate a new value for the newly created field.

     Example: ([ALL_CL14] &" "& [LUCLW_CL10] &" "& [LULCL_CL10])

   This new field provides you with a single numeric value (in a text format) that collectively represents the combined cluster numbers from the selected watershed landscape cluster, the watershed land cover cluster, and the local land cover cluster.

3. The stream network data for this project includes streams that fall outside of the state of Missouri. In order to properly use this data, the user must select only

those streams that fall within the state and also none of their watershed occurring outside of the state. This is necessary because landscape variables were only summarized for within the boundary of Missouri and any stream segment outside of the state or that has a portion of its watershed outside of the state will have inaccurate landscape or land cover data.

- Select in-state streams using the OUT and OUT_ABOVE attributes.
- Query for "OUT = 0 AND OUT_ABOVE = 0" to obtain stream segments that fall within Missouri and have their entire watershed within Missouri.
- The result should be 170,345 segments from the overall 217,860 segments.

4. Next, select from the above set the range of stream size you are interested in.
   - Example, for headwater streams select values ranging from 1 to 10 from the LINK attribute.

5. Identify which stream segments have headwater impoundments
   - Both the CLUST_FULL and CLUST_RED coverages have 8 attributes that can be used to identify stream segments with and without impoundments. The user must select one set of two (e.g., 1a and 1b) in order to separate those segments with and without headwater impoundments within their watersheds.

1a	HWP = Water bodies 2-15 acres that intersect stream segments (0, no water body; 1, water body present)
1b	HWP_ABOVE = Number of water bodies 2-15 acres above segment
2a	HWPT = Water bodies 2-15 acres that fall within a 100m tolerance of stream segments (0, no water body; 1, water body present)
2b	HWPT_ABOVE = Number of water bodies 2-15 acres and within tolerance above segment
3a	HWP2 = Water bodies 2-50 acres that intersect stream segments (0, no water body; 1, water body present)
3b	HWP2_ABOVE = Number of water bodies 2-50 acres above segment
4a	HWP2T = Water bodies 2-50 acres that fall within a 100m tolerance of stream segments (0, no water body; 1, water body present)
4b	HWP2T_ABOVE = Number of water bodies 2-50 acres and within tolerance above segment

6. Next, create 2 separate shapefiles for comparison.
    - Example: Say you selected 2a and 2b in step 5.
        - A shapefile of streams **with** headwater impoundments in their watershed
            - Query the "HWPT = 1" to identify streams intersecting headwater impoundments
            - Query for "HWPT _ABOVE > 0" to identify streams with headwater impoundments in their watersheds
            - Save combined query results to a new shapefile
                - e.g., LK1_HP1.SHP represents streams Link 1-10 with headwater impoundments
        - A shapefile of streams **without** headwater impoundments in their watershed
            - Query for "HWPT = 0" to identify streams not intersecting headwater impoundments
            - Query for "HWPT_ABOVE = 0" to identify streams with no impoundments in their watersheds
            - Save combined query results to a new shapefile
                - e.g., LK1_HP0.SHP represents streams Link 1-10 with no headwater impoundments

7. For each new shapefile, open and then SUMMARIZE the .DBF table by the concatenated cluster values created in Step 2 above (e.g., ALL_CAT14)
    - Open the shapefile table, go to table properties, and select SUMMARIZE
    - Name the resulting summary table according to the number of clusters from the selected landscape dataset, the stream size range, and whether or not it is for streams with or without headwater impoundments (e.g., the summary table for ALL_CAT14, for stream links 1-10, with no headwater impoundments might be named ALL14_CAT_SUM_LK1to10_HP0.DBF)
    - Sort summary table by descending value
        - This table is used to assess the replication potential for the chosen cluster. Concatenated cluster values with larger numbers of stream segments have a greater potential for replication.
        - Note: the number of potential replicates in the summary table for the concatenated cluster values is misleading. This number represents the total number of stream *segments* within each class, not the number of streams. In many instances several stream segments from the same stream are represented in the selection. Consequently, the only way to truly assess replication potential and select specific study sites is to visually examine the results within a GIS (See Step 10).

8. Use the cluster means tables to find suitable cluster numbers and their combinations to determine which values in the concatenated cluster field are suited to assessing the potential effects of headwater impoundments. Refer to question #3 discussed above.

9. Next, for each shapefile (with and without headwater impoundments), query for each of the desired combinations of cluster numbers from the new Concatenated Value Field created in step 2.
   - For example, if you decide that cluster number 14 from the Full Set, 4 from the Watershed Land Cover Set, and 7 from the Local Land Cover Set were appropriate clusters for research then you would query for "14 4 7" from the Concatenated Value Field to find stream segments that share this combination of cluster numbers.
   - In each instance convert the selection to a new shapefile and give it an informative name. For instance, ALL14_LK1to10_1447_HP0 for the shapefile containing segments with no headwater impoundments and ALL14_LK1to10_1447_HP1 for the shapefile containing segments with headwater impoundments.
   - The above steps will have to be carried out for each of the selected concatenated values.

10. Simultaneously display each combination of the corresponding shapefiles developed in Step 9 to visually examine the results. Give those segments with headwater impoundments one color or size scheme and those without another scheme.

11. Finally, select the initial pool of potential replicate sites by creating a new field in each shapefile (e.g., POT_REPS) and give each selected site a value of 1.

Because the geospatial datasets used to generate the cluster groupings for this project are relatively coarse scale (1:100,000, 1:250,000, or 1:500,000) and are not without error, the above process should only be used as an efficient initial coarse-screening of potential study designs and site-selection tool. Whenever possible, in Step 11 the potential replicates should be selected so that they are geographically situated as close together as possible. In most instances, this will increase the probability that the selected sites are similar with regard to the classification variables. Also, additional higher resolution datasets (e.g., geology, soils, Digital Ortho Quads) should be used along with field visits to further assess the relative similarity of the initial pool of potential replicates.

Task 9: Assess opportunities or limitations for treatment vs. control and correlative experimental designs for three stream size classes

Task 9 Methods

If you have read the information provided under Task 8 you are quite aware of the fact that there are an amazing number of options for the end user to help develop a study to assess the potential beneficial or negative effects of headwater impoundments. A different decision for any of the six questions that must be answered will alter the outcome. Also, even after the questions are answered there are numerous stream types (i.e., combinations of watershed landscape, watershed land cover, and local land cover) that could serve as a potential focus for research. It was impossible for us to assess replication potential and study design options for all of the possible outcomes. However, we did want to provide MDNR and the advisory committee with a reasonable assessment of opportunities and limitations for devising field studies using the GIS databases we created.

Specifically, we wanted to assess possible differences in replication potential and study design options;

1. as stream size changed
2. between the cluster sets produced by the Full and Reduced Set of landscape variables, and
3. as the number of clusters within each set changed

To assess any possible differences we answered the six questions discussed under Task 8 in the following manner.

1. We selected both the Full and Reduced Set of watershed landscape variables. In conjunction with each of these sets we decided to use both the Watershed and Local Land Cover Sets.
2. Using the diagnostic plots we decided to compare the 14, 16, and 18 cluster results for the Full Set and 12, 14, and 16 cluster results for the Reduced Set. In both instances we concatenated the cluster numbers for each of these cluster sets with the cluster numbers from the 10 cluster results for both the Watershed and Local Land Cover Sets.
3. Using the cluster means tables we selected specific cluster numbers that were relatively homogenous in terms of watershed landscape character for the Full and Reduced Sets (e.g., for the Full set we selected cluster number 14 for both the 14 and 16 cluster sets and cluster number 10 from the 18 cluster set). From the watershed land cover cluster means table (10 clusters) we selected cluster numbers 4 and 9, which are both a mixture cropland and grassland (See Table 11). From the local land cover cluster means table we selected cluster numbers 3, 6, and 7, which were also primarily a mixture cropland and grassland (See Table 14). The reason we selected these specific clusters for the land cover sets

was because the cropland/grassland mixture is the dominant land cover combination within the overall study area.

4. We selected three stream sizes to examine based on ranges of Shreve link. For discussion purposes we simply defined these size classes as headwaters (Link = 1-10), creeks (Link 50-100), and small rivers (Link 500-1000).
5. We defined headwater impoundments as those water bodies that were 2-15 acres in size.
6. We also chose to use the 100-meter tolerance for identifying stream segments that intersected headwater impoundments.

Task 9 Results

Table 17 provides the results of our assessment. For each landscape dataset we examined 6 “stream types”, two within each of the three cluster sets. For the headwater size class the results were simply broken into two categories to show the number of stream segments with and without headwater impoundments for each stream type (i.e., concatenated cluster number). For the two larger stream size classes we broke the results into four categories. Specifically, for creeks the table shows the number of segments without headwater impoundments and the number of segments with greater than 1, less than 10, and between 10 and 50 impoundments within the watershed. Finally, for small rivers the table shows the number of segments without headwater impoundments and the number of segments with less than 20, between 20 and 100, and greater than 100 impoundments within their watershed.

Differences as Stream Size Changed

In every instance there appears to be sufficient replication potential for treatment vs. control experiments with the headwater size class. In most instances there are hundreds of segments with and thousands of segments without headwater impoundments to select from for designing a study (Table 17, Figures 23, 24). As stream size increases, however, the potential for devising treatment vs. control studies diminishes quickly (Table 17). As we defined creeks (link = 50 to 100), there is still some potential for devising such studies (Figure 25). This assumes, however, that additional scrutiny as to the relative similarity of the segments does not find the relatively few segments (range 3-45) without headwater impoundments to be unsuitable candidates for research. Users could certainly decrease the range of values (e.g., 20 to 50) to assess potential cumulative effects and this would certainly increase the replication potential, but whether this would provide enough segments for a treatment vs. control study is uncertain. In no instance, was there an opportunity for devising a treatment vs. control study for the largest size class we examined (link = 500 to 1000). All segments within this size range had at least 10 headwater impoundments within their watersheds. Consequently, for larger streams a correlative approach would have to be taken where differences in selected environmental variables are assessed as the number of headwater impoundments within the watershed changes. In most instances the potential replicates in the largest size class provided a good range of values suited to a correlative study design (e.g., 20 to 113, Figure 26).

Table 17. Results of an assessment of replication potential and study design options.

			Stream Size Categories								
			Link 1-10		Link 50-100				Link 500-1000		
Landscape Dataset	Number of Clusters	Concatenated Cluster Number	without impoundments	with impoudments	without impoundments	with impoundments on and/or above	<10 impoundments above	10-50 impoundments above	<20 impoundments above	20-100 impoundments above	>100 impoundments above
Full	**14**										
		14 4 7	10457	694	45	79	79	0	4	0	0
		14 9 3	520	96	16	144	97	47	1	118	24
	16										
		14 9 6	7047	666	21	97	73	24	9	24	5
		14 9 3	404	68	10	77	59	18	4	60	12
	18										
		10 9 6	5633	551	24	77	70	7	5	18	5
		10 4 7	6348	409	21	66	66	0	7	0	0
Reduced	**12**										
		9 9 6	8187	870	36	158	123	35	9	32	12
		1 9 6	3400	360	12	61	53	8	16	39	11
	14										
		10 9 6	790	103	3	63	45	18	27	39	14
		13 9 6	7923	835	36	153	118	35	9	31	10
	16										
		15 9 3	451	122	18	106	77	29	4	101	15
		15 9 6	7745	817	36	148	116	32	9	26	10

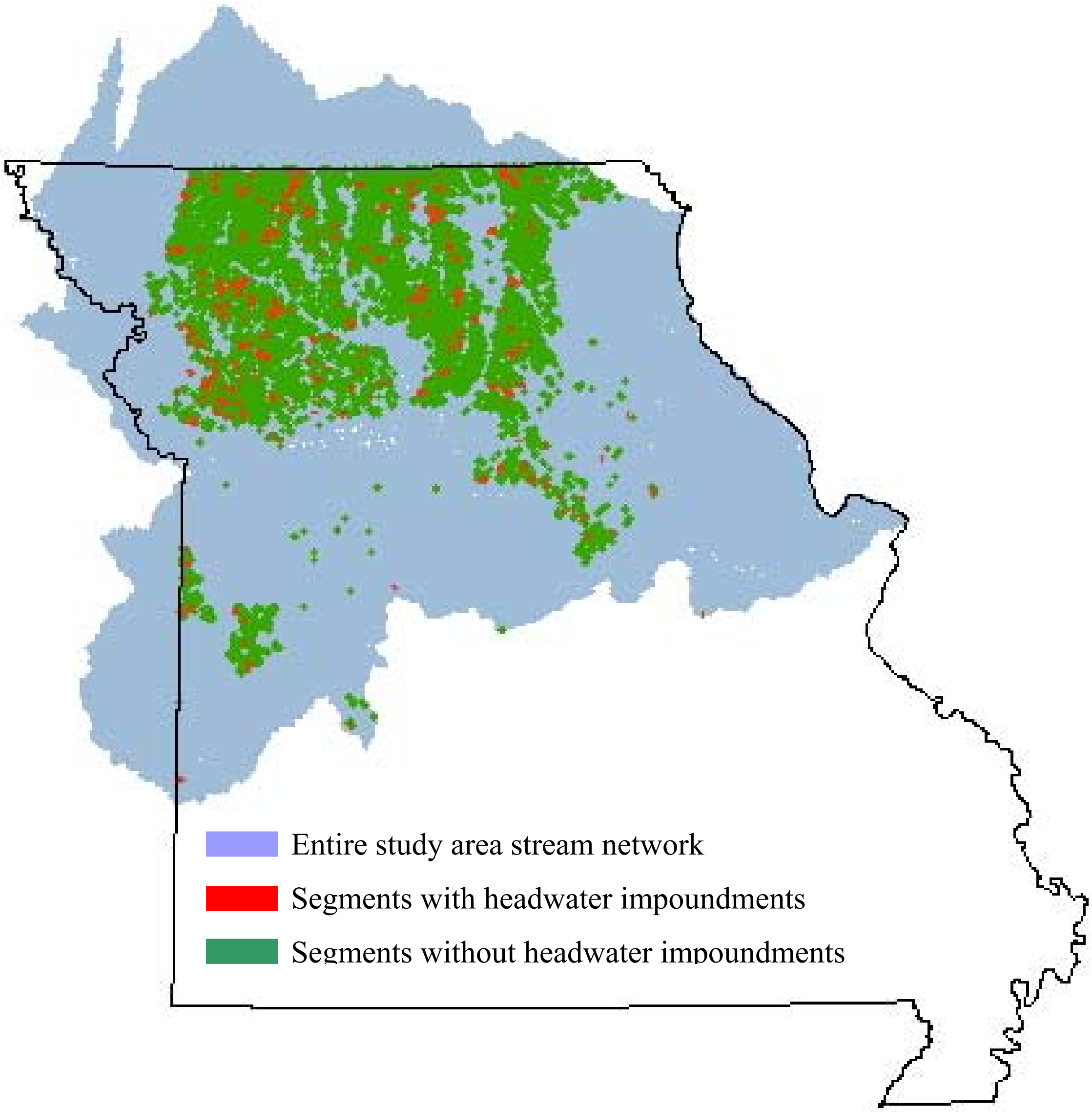


Figure 23. Locations of stream segments with and without headwater impoundments for the concatenated cluster number “14 4 7”, derived from the 14 cluster set for the Full Set of landscape variables and the 10 cluster sets for both watershed and local land cover.

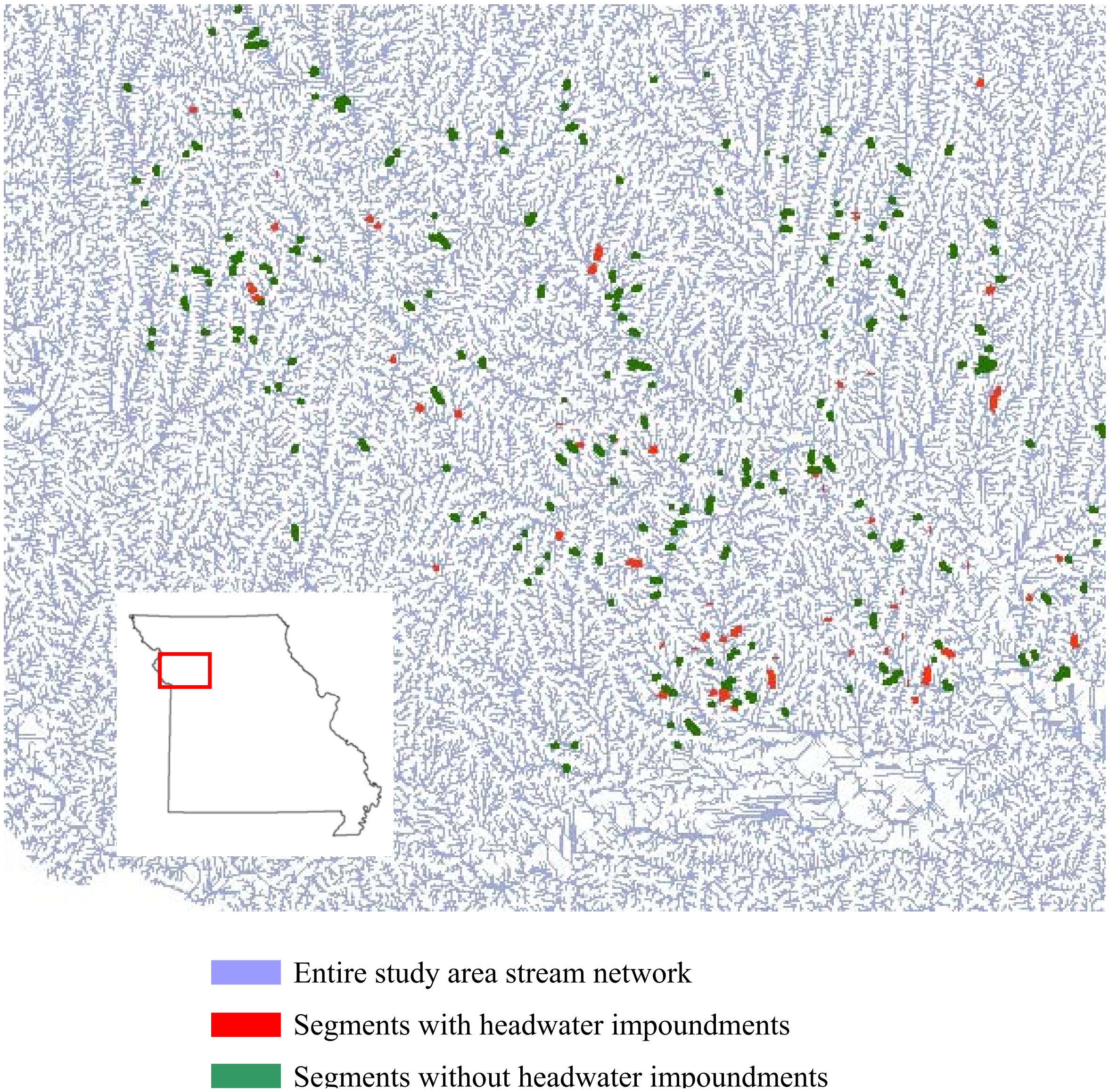


Figure 24. Locations of headwater (link = 1 to 10) stream segments with and without headwater impoundments for the concatenated cluster number “15 9 3”, derived from the 16 cluster set for the Reduced Set of landscape variables and the 10 cluster sets for both watershed and local land cover.

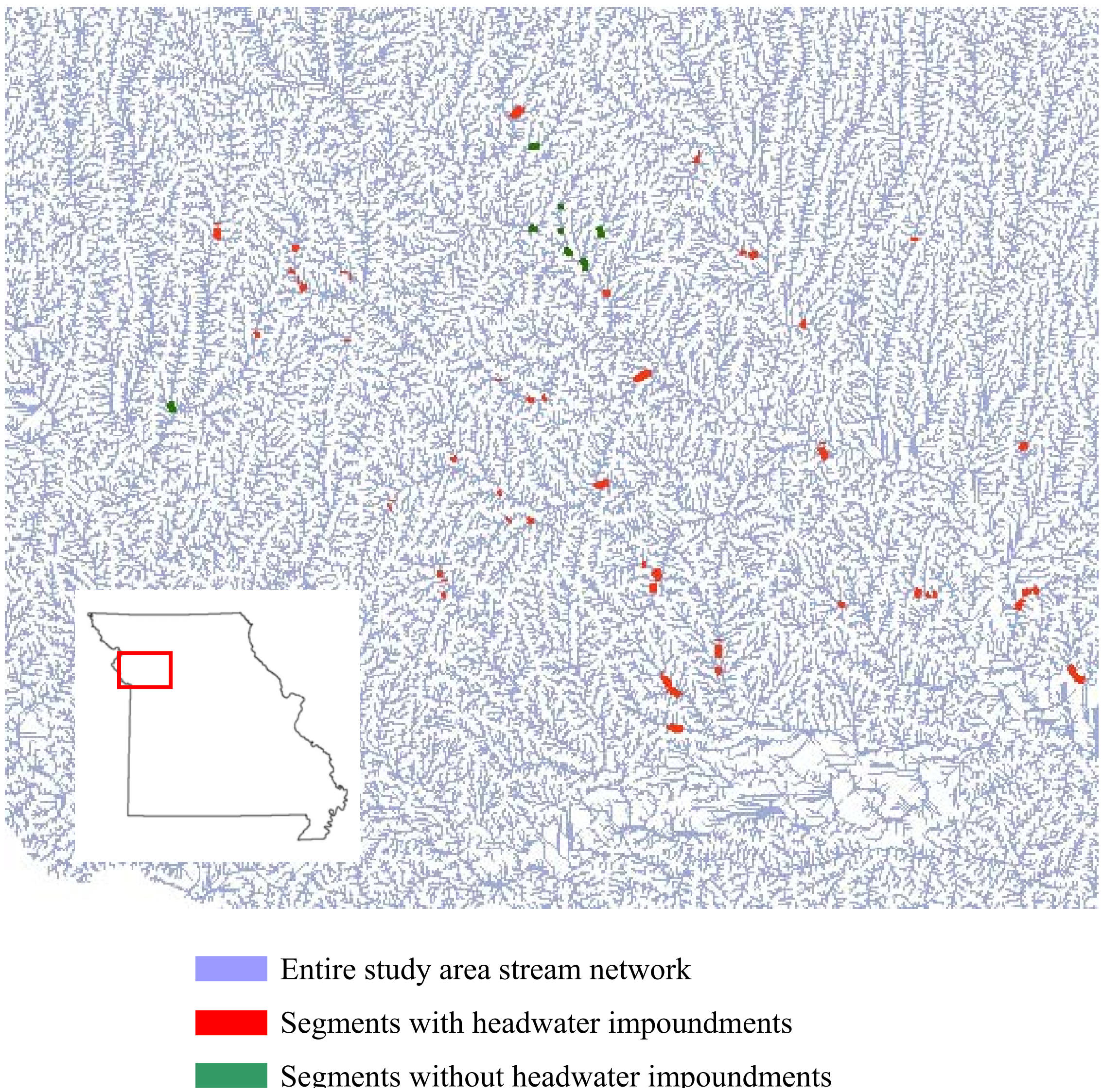


Figure 25. Locations of creek (link = 50 to 100) stream segments with and without headwater impoundments for the concatenated cluster number “15 9 3”, derived from the 16 cluster set for the Reduced Set of landscape variables and the 10 cluster sets for both watershed and local land cover. Compare with Figure 23 to see how the number of segments without headwater impoundments dramatically decreases. Also, notice how the 18 stream segments listed in Table 17, actually only represent 4 different streams.

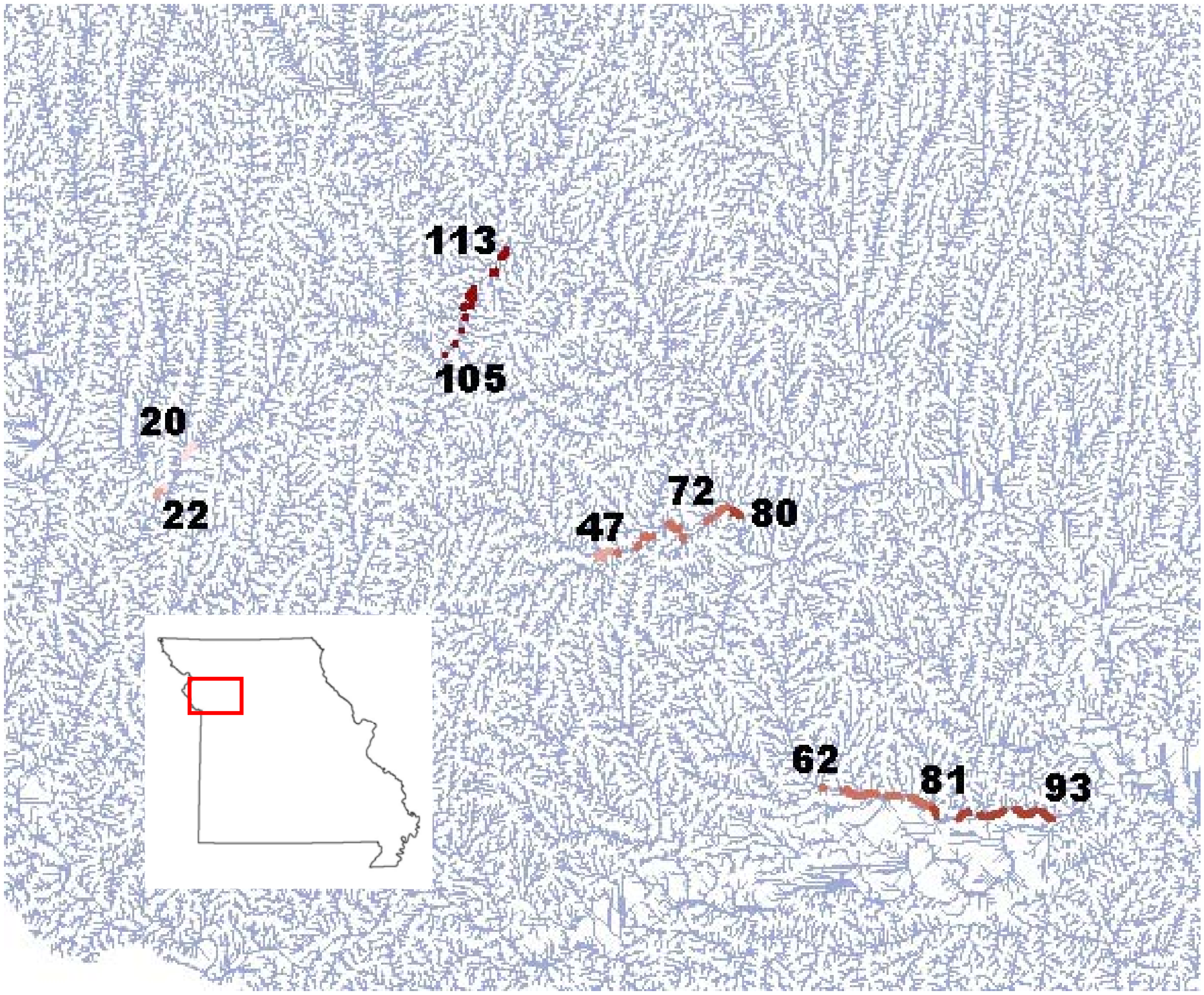


Figure 26. Locations of small river (link = 500 to 1000) stream segments for the concatenated cluster number "15 9 3", derived from the 16 cluster set for the Reduced Set of landscape variables and the 10 cluster sets for both watershed and local land cover. All segments are color-coded (darker colors = more headwater impoundments) and select segments are labeled according to the actual number of headwater impoundments within the watershed of that segment. No segments for this size class could be found without headwater impoundments within their watersheds. The range and relatively even spread of values suggests that the opportunity for a relatively sound correlative design does exist for larger streams.

Differences between Full and Reduced Sets

There were surprisingly few differences in the replication potential between the Full and Reduced Sets. This is especially true for the two smallest size classes. For the largest size class, however, the Reduced Set appears to provide a more even distribution in the number of segments within each headwater impoundment category. One major advantage of the Reduced Set is that it is easier to find clusters that are relatively "homogenous" in terms of landscape character, which increases the number of options available (i.e., stream types) for designing a study (Table 18). This certainly makes sense considering the fact that only 8 variables were used to generate the clusters for the Reduced Set, while 19 variables were used for the Full Set.

Differences as the Number of Clusters is Increased

We expected that the replication potential would dramatically decrease as we increased the number of clusters in both the Full and Reduced Sets. Yet, at least for the 12 stream types that we examined, there appears to be little change in replication potential. Overall, the number of potential replicates does show a decline as you increase the number of clusters, yet there are also instances where replication potential actually increases with the number of clusters.

Table 18. Assessment of the relative homogeneity of clusters generated from the Full and Reduced Sets of landscape variables. For a cluster to be counted in the third column it had to have greater than 70% its watershed in just a single landscape category across all of the landscape factors. For instance, for the Full Set a cluster was counted only if the cluster means were >70% in one system geology class, >70% in one relief class, >70% in one surface texture class, and >70% in one hydrologic soil group. Only one third or less of the clusters in the Full Set meet this cutoff, while 60 to 83% of the clusters for the Reduced Set meet the cutoff.

Landscape Dataset	Number of Clusters	Number of Clusters with >70% of their watershed in one landscape category across all landscape factors	Percent of Clusters
Full Set	14	2	14.3
	16	5	31.3
	18	6	33.3
Reduced Set	12	10	83.3
	14	9	64.3
	16	10	62.5

Literature Cited

Anderberg, M. R. 1973. Cluster Analysis for Applications. Academic Press, New York, NY.

Bock, H. H. 1985. On some significance tests in cluster analysis. Journal of Classification, 2, 77-108.

Calinski, T. and J. Harabasz. 1974.   A dendrite method for cluster analysis. Communications in Statistics 3: 1-27.

Cooper, M. C. and G. W. Milligan. 1988. The effect of error on determining the number of clusters. Pages 319-328, In: The Proceedings of the International Workshop on Data Analysis, Decision Support and Expert Knowledge Representation in Marketing and Related Areas of Research.

Doisy, K. E. and C. F. Rabeni. 2004. A Literature Review on the Effects of PL-566 Headwater Impoundments on the Ecological Integrity of Missouri Stream Systems. Final report to the Missouri Department of Natural Resources under agreement MDNR G02-566-1. 55 pp.

Everitt, B. S. 1979. Unresolved problems in cluster analysis. Biometrics, 35, 169-181.

Greenlee, D. D. 1987. Raster and vector processing for scanned linework. Photogrammetric Engineering and Remote Sensing. 53(10): 1383-1387.

Hack, J.T. 1957. Studies of longitudinal stream profiles in Virginia and Maryland. U.S. Geological Survey Professional Paper 294B: 45-97.

Hansen, W. F. 2001. Identifying stream types and management implications. Forest Ecology and Management 143: 39-46.

Hartigan, J. A. 1985. Statistical theory in clustering. Journal of Classification 2: 63-76.

Horton, R. E., 1945. Erosional development of streams and their drainage basins, hydrophysical approach to quantitative morphology. Geological Society of America Bulletin 56: 275-370.

Hynes, H. B. N. 1970. The Ecology of Running Waters. University of Toronto Press.

Hynes, H. B. N. 1975. The stream and its valley. Verh. Int. Theor. Ang. Limnol. 19: 1-15.

Jenson, S. K. and J. O. Domingue. 1988. Extracting topographic structure from digital elevation data for geographic information system analysis. Photogrammetric Engineering and Remote Sensing 54(11): 1593-1600.

Jongman, R. H. G., ter Braak, C. J. F. and van Tongeren, O. F. R. 1995. Data Analysis in Community and Landscape Ecology. Cambridge University Press, Cambridge, UK.

Knighton, D. 1998. Fluvial forms and processes: a new perspective. Oxford University Press. New York, NY.

Lammert, M. and J. D. Allan, 1999. Assesing biological integretity of streams: effects of scale in measuring the influence of land use/cover and habitat structure on fish and invertebrates. Environmental Management 23: 257-270.

Lamphear, D.W. and J. Lewis. 1994. Stream_o.aml. USDA Redwood Sciences Laboratory, Arcata, CA.

MDNR. 1992. Geologic Map of Missouri. Available online at: http://msdisweb.missouri.edu/datasearch/metadata/utm/st_geol_utm.xml

Milligan, G. W. and M. C. Cooper. 1985. An examination of procedures for determining the number of clusters in a data set. Psychometrika 50: 159-179.

Nigh, T. A. and W. A. Schroeder. 2002. Atlas of Missouri Ecoregions. Missouri Department of Conservation, Jefferson City, MO.

Nino, Y. 2002. Simple model for downstream variation of median sediment size in Chilean Rivers. Journal of Hydraulic Engineering 128: 934-941.

NRCS. 1999. Dams in Danger. Available online at: http://www.nrcs.usda.gov/programs/ws_reinvent/dams_in_danger/20_DAMS/Missouri.pdf

Panfil, M. S. and R. B. Jacobson. 2001. Relations among geology, physiography, land use, and stream habitat conditions in the Buffalo and Current River systems, Missouri and Arkansas. U.S. Geological Survey, Biological Science Report, USGS/BRD/BSR-2001-005.

Pflieger, W. L. 1997. The Fishes of Missouri, 2nd edition. Missouri Department of Conservation, Jefferson City, MO.

Richards, C., L. B. Johnson, and G. E. Host. 1996. Landscape-scale influences on stream habitat and biota. Canadian Journal of Fisheries and Aquatic Science 53 (Suppl. 1): 295-311.

Salvador, S. and P. Chan. 2003. Determining the number of clusters/segments in hierarchical clustering/segmentation algorithms. Department of Computer Sciences Technical Report CS-2003-18, Florida Institute of Technology. Melbourne, FL.

Sarle, W. S. 1983. The Cubic Clustering Criterion. SAS Technical Report A-108, Cary, NC: SAS Institute Inc.

SAS Institute. 2001. SAS User's Guide: Statistics. Version 8.2. Cary, NC, SAS Institute Inc.

Tarboton, D. G., R. L. Bras, I. Rodriguez-Iturbe. 1991. On the extraction of channel networks from digital elevation data. Hydrological Processes 5: 81-100.

USDA. 1985. Big Creek-Hurrican Creek Watershed: Watershed Plan – Environmental Impact Statement. USDA Soil Conservation Service. Columbia, Missouri.

Vandike, J. E. 1995. Missouri State Water Plan Series; Volume 1: Surface Water Resources of Missouri. Missouri Department of Natural Resources, Division of Geology and Land Survey, Water Resources Report Number 45.

Vannote, R. L., G. W. Minshall, K. W. Cummins, J. R. Sedell, and C. E. Cushing. 1980. The river continuum concept. Canadian Journal of Fisheries and Aquatic Science 37: 130-137.

Wang, L., J. Lyons, P. Rasmussen, P. Seelbach, T. Simon, M. Wiley, P. Kanehl, E. Baker, S. Niemela, and P. M. Stewart. 2003. Watershed, reach, and riparian influences on stream fish assemblages in the Nothern Lakes an Forest Ecoregion, USA. Canadian Journal of Fisheries and Aquatic Science 60: 491-505.

Appendix A. Attribute descriptions for the raw watershed data generated for each individual stream segment. Corresponds with the NETW_ALL coverage.

Attribute	Description
FNODE#	Internal node number for the beginning of an arc (from-node)
TNODE#	Internal node number for the end of an arc (to-node)
LPOLY#	Internal node number for the left polygon
RPOLY#	Internal node number for the right polygon
LENGTH	Length (in meters) of stream segment
TNETW_ALL#	Internal feature number. Sequential unique whole numbers that are automatically generated.
TNETW_ALL-ID	User defined feature number
ARCIDNUM	Stream Segment Number (Unique Identifier used to link all tables and coverages)
HECTARES2	Area of individual stream segmentshed in hectares
SHED_INCLUS	Area of watershed in hectares
STRAHLER	Strahler Stream Order
LINK	Shreve Link
LULCW_CROP	Percent of watershed classified as Cropland
LULCW_FOR	Percent of watershed classified as Forest
LULCW_GRASS	Percent of watershed classified as Grassland
LULCW_SWAMP	Percent of watershed classified as Swamp
LULCW_URB	Percent of watershed classified as Urban
LULCW_WAT	Percent of watershed classified as Water
GEO1W_ALUV	Percent of watershed classified General Geology Type Alluvium
GEO1W_CLAY	Percent of watershed classified General Geology Type Clay
GEO1W_DOL	Percent of watershed classified General Geology Type Dolomite
GEO1W_LIME	Percent of watershed classified General Geology Type Limestone
GEO1W_SAND	Percent of watershed classified General Geology Type Sand
GEO2W_DEV	Percent of watershed classified System Geology Type Devonian
GEO2W_MSPI	Percent of watershed classified System Geology Type Mississippian
GEO2W_ORD	Percent of watershed classified System Geology Type Ordovician
GEO2W_PENN	Percent of watershed classified System Geology Type Pennsylvanian
GEO2W_QUAT	Percent of watershed classified System Geology Type Quaternary
GEO2W_SILR	Percent of watershed classified System Geology Type Silurian
GEO3W_ALN	Percent of watershed classified Series Geology Type Alexandrian/Niagar
GEO3W_ATDES	Percent of watershed classified Series Geology Type Atokan/Desmoinesia
GEO3W_CAN	Percent of watershed classified Series Geology Type Candadian
GEO3W_CHCH	Percent of watershed classified Series Geology Type Cinncinnatian/Champlanian
GEO3W_CHMP	Percent of watershed classified Series Geology Type Champlanian
GEO3W_DES	Percent of watershed classified Series Geology Type Desmoinesia
GEO3W_HOL	Percent of watershed classified Series Geology Type Holocene
GEO3W_KIND	Percent of watershed classified Series Geology Type Kinderhookian
GEO3W_LOMI	Percent of watershed classified Series Geology Type Alexandrian/Niagar
GEO3W_MER	Percent of watershed classified Series Geology Type Meramecian

Appendix A. Continued.

GEO3W_MISS	Percent of watershed classified Series Geology Type Missourian
GEO3W_OSAG	Percent of watershed classified Series Geology Type Osagean
GEO3W_VIRG	Percent of watershed classified Series Geology Type Virgilian
HGW_B	Percent of watershed classified as Hydrologic Soil Group B in STATSGO dataset
HGW_C	Percent of watershed classified as Hydrologic Soil Group C in STATSGO dataset
HGW_D	Percent of watershed classified as Hydrologic Soil Group D in STATSGO dataset
SFW_CL	Percent of watershed classified as Clay Loam in STATSGO dataset
SFW_CRSIL	Percent of watershed classified as Cherty/Silty Loam in STATSGO dataset
SFW_CRVSL	Percent of watershed classified as Very Cherty/Silty Loam in STATSGO dataset
SFW_FSL	Percent of watershed classified as Fine Sandy Loam in STATSGO dataset
SFW_L	Percent of watershed classified as Cherty/Silty Loam in STATSGO dataset
SFW_LS	Percent of watershed classified as Loam in STATSGO dataset
SFW_SIC	Percent of watershed classified as Silty Clay in STATSGO dataset
SFW_SICL	Percent of watershed classified as Silty Clay Loam in STATSGO dataset
SFW_SIL	Percent of watershed classified as Cherty/Silty Loam in STATSGO dataset
SFW_STL	Percent of watershed classified as Stony Loam in STATSGO dataset
SFW_VAR	Percent of watershed classified as Variable in STATSGO dataset
SFW_STSIL	Percent of watershed classified as Stony Silt Loam in STATSGO dataset
SF2W_CHRT	Percent of watershed classified as Cherty (combination of STATSGO CRSIL, CRVSL)
SF2W_CLAY	Percent of watershed classified as Clays (combination of STATSGO CL, SIC, SICL)
SF2W_LOAM	Percent of watershed classified as Loams (combination of STATSGO L, SIL, VAR)
SF2W_SAND	Percent of watershed classified as Sandy (combination of STATSGO FSL, LS)
SF2W_STONY	Percent of watershed classified as Stony (combination of STATSGO STL, STSIL)
RELCATW_1	Percent of watershed in Relief Category 1 (0-50 feet)
RELCATW_2	Percent of watershed in Relief Category 2 (51-100 feet)
RELCATW_3	Percent of watershed in Relief Category 3 (101-200 feet)
RELCATW_4	Percent of watershed in Relief Category 4 (201-300 feet)
RELCATW_5	Percent of watershed in Relief Category 5 (301-500 feet)
RELCATW_6	Percent of watershed in Relief Category 6 (501-700 feet)
SLOP_WAT	Mean slope of watershed
OUT	Identifies stream segements that fall outside of Missouri and will not have landscape attributes
OUT_ABOVE	Summarizes the number of streams above each segment that are outside of Missouri
HW_POND	Indicates presence/absence of an impoundment (size 2-15 acres) on a stream segment
HWP_ABOVE	Number of impoundments (size 2-15 acres) above segment
HWPT	Indicates presence/absence of impoundment (size 2-15 acres) within a 100 meter tolerance
HWPT_ABOVE	Number of impoudments (size 2-15 acres) within a 100 meter tolerance above each segment
HWP2	Indicates presence/absence of an impoundment (size 2-50 acres) on a stream segment
HWP2_ABOVE	Number of impoundments (size 2-50 acres) above segment
HWP2T	Indicates presence/absence of impoundment (size 2-50 acres) within a 100 meter tolerance
HWP2T_ABOVE	Number of impoudments (size 2-50 acres) within a 100 meter tolerance above each segment
WB_INT	Indicates presence/absence of any size waterbody intersecting the stream segment

Appendix B. Attribute descriptions for the raw segmentshed data generated for each individual stream segment. Corresponds with the NETL_ALL coverage.

Attribute	Description
FNODE#	Internal node number for the beginning of an arc (from-node)
TNODE#	Internal node number for the end of an arc (to-node)
LPOLY#	Internal node number for the left polygon
RPOLY#	Internal node number for the right polygon
LENGTH	Length (in meters) of stream segment
TNET_ALL#	Internal feature number. Sequential unique whole numbers that are automatically generated.
TNET_ALL-ID	User defined feature number
ARCIDNUM	Stream Segment Number (Unique Identifier used to link all tables and coverages)
HECTARES2	Area of individual stream segmentshed in hectares
SHED_INCLUS	Area of watershed in hectares
STRAHLER	Strahler Stream Order
LINK	Shreve Link
LULCL_CROP	Percent of local segmentshed classified as Cropland
LULCL_FOR	Percent of local segmentshed classified as Forest
LULCL_GRASS	Percent of local segmentshed classified as Grassland
LULCL_SWAMP	Percent of local segmentshed classified as Swamp
LULCL_URB	Percent of local segmentshed classified as Urban
LULCL_WAT	Percent of local segmentshed classified as Water
GEO1L_ALUV	Percent of local segmentshed classified General Geology Type Alluvium
GEO1L_CLAY	Percent of local segmentshed classified General Geology Type Clay
GEO1L_DOL	Percent of local segmentshed classified General Geology Type Dolomite
GEO1L_LIME	Percent of local segmentshed classified General Geology Type Limestone
GEO1L_SAND	Percent of local segmentshed classified General Geology Type Sand
GEO2L_DEV	Percent of local segmentshed classified System Geology Type Devonian
GEO2L_MSPI	Percent of local segmentshed classified System Geology Type Mississippian
GEO2L_ORD	Percent of local segmentshed classified System Geology Type Ordovician
GEO2L_PENN	Percent of local segmentshed classified System Geology Type Pennsylvanian
GEO2L_QUAT	Percent of local segmentshed classified System Geology Type Quaternary
GEO2L_SILR	Percent of local segmentshed classified System Geology Type Silurian
GEO3L_ALN	Percent of local segmentshed classified Series Geology Type Alexandrian/Niagar
GEO3L_ATDES	Percent of local segmentshed classified Series Geology Type Atokan/Desmoinesia
GEO3L_CAN	Percent of local segmentshed classified Series Geology Type Candadian
GEO3L_CHCH	Percent of local segmentshed classified Series Geology Type Cinncinnatian/Champlanian
GEO3L_CHMP	Percent of local segmentshed classified Series Geology Type Champlanian
GEO3L_DES	Percent of local segmentshed classified Series Geology Type Desmoinesia
GEO3L_HOL	Percent of local segmentshed classified Series Geology Type Holocene
GEO3L_KIND	Percent of local segmentshed classified Series Geology Type Kinderhookian
GEO3L_LOMI	Percent of local segmentshed classified Series Geology Type Alexandrian/Niagar
GEO3L_MER	Percent of local segmentshed classified Series Geology Type Meramecian
GEO3L_MISS	Percent of local segmentshed classified Series Geology Type Missourian
GEO3L_OSAG	Percent of local segmentshed classified Series Geology Type Osagean

Appendix B. Continued.

GEO3L_VIRG	Percent of local segmentshed classified Series Geology Type Virgilian
HGL_B	Percent of local segmentshed classified as Hydrologic Soil Group B in STATSGO dataset
HGL_C	Percent of local segmentshed classified as Hydrologic Soil Group C in STATSGO dataset
HGL_D	Percent of local segmentshed classified as Hydrologic Soil Group D in STATSGO dataset
SFL_CL	Percent of local segmentshed classified as Clay Loam in STATSGO dataset
SFL_CRSIL	Percent of local segmentshed classified as Cherty/Silty Loam in STATSGO dataset
SFL_CRVSL	Percent of local segmentshed classified as Very Cherty/Silty Loam in STATSGO dataset
SFL_FSL	Percent of local segmentshed classified as Fine Sandy Loam in STATSGO dataset
SFL_L	Percent of local segmentshed classified as Cherty/Silty Loam in STATSGO dataset
SFL_LS	Percent of local segmentshed classified as Loam in STATSGO dataset
SFL_SIC	Percent of local segmentshed classified as Silty Clay in STATSGO dataset
SFL_SICL	Percent of local segmentshed classified as Silty Clay Loam in STATSGO dataset
SFL_SIL	Percent of local segmentshed classified as Cherty/Silty Loam in STATSGO dataset
SFL_STL	Percent of local segmentshed classified as Stony Loam in STATSGO dataset
SFL_VAR	Percent of local segmentshed classified as Variable in STATSGO dataset
SFL_STSIL	Percent of local segmentshed classified as Stony Silt Loam in STATSGO dataset
SF2L_CHRT	Percent of local segmentshed classified as Cherty (combination of STATSGO CRSIL, CRVSL)
SF2L_CLAY	Percent of local segmentshed classified as Clays (combination of STATSGO CL, SIC, SICL)
SF2L_LOAM	Percent of local segmentshed classified as Loams (combination of STATSGO L, SIL, VAR)
SF2L_SAND	Percent of local segmentshed classified as Sandy (combination of STATSGO FSL, LS)
SF2L_STONY	Percent of local segmentshed classified as Stony (combination of STATSGO STL, STSIL)
RELCATL_1	Percent of local segmentshed in Relief Category 1 (0-50 feet)
RELCATL_2	Percent of local segmentshed in Relief Category 2 (51-100 feet)
RELCATL_3	Percent of local segmentshed in Relief Category 3 (101-200 feet)
RELCATL_4	Percent of local segmentshed in Relief Category 4 (201-300 feet)
RELCATL_5	Percent of local segmentshed in Relief Category 5 (301-500 feet)
RELCATL_6	Percent of local segmentshed in Relief Category 6 (501-700 feet)
RCH_GRAD	Stream segment Gradient
MAX_ELEV	Maximum elevation within local segementshed
MIN_ELEV	Minimum elevation within local segementshed
MN_SLOPE	Mean slope of local segmentshed
OUT	Identifies stream segements that fall outside of Missouri and will not have landscape attributes
OUT_ABOVE	Summarizes the number of streams above each segment that are outside of Missouri
HW_POND	Indicates presence/absence of an impoundment (size 2-15 acres) on a stream segment
HWP_ABOVE	Number of impoundments (size 2-15 acres) above segment
HWPT	Indicates presence/absence of impoundment (size 2-15 acres) within a 100 meter tolerance
HWPT_ABOVE	Number of impoudments (size 2-15 acres) within a 100 meter tolerance above each segment
HWP2	Indicates presence/absence of an impoundment (size 2-50 acres) on a stream segment
HWP2_ABOVE	Number of impoundments (size 2-50 acres) above segment
HWP2T	Indicates presence/absence of impoundment (size 2-50 acres) within a 100 meter tolerance
HWP2T_ABOVE	Number of impoudments (size 2-50 acres) within a 100 meter tolerance above each segment
WB_INT	Indicates presence/absence of any size waterbody intersecting the stream segment

Appendix C. Attribute descriptions for the Full Set of watershed landscape variables used in the cluster analyses. Corresponds with the CLUSTALL_VARS coverage.

Attribute	Description
FNODE#	Internal node number for the beginning of an arc (from-node)
TNODE#	Internal node number for the end of an arc (to-node)
LPOLY#	Internal node number for the left polygon
RPOLY#	Internal node number for the right polygon
LENGTH	Length (in meters) of stream segment
CLUSTALL_VARS#	Internal feature number. Sequential unique whole numbers that are automatically generated.
CLUSTALL_VARS-ID	User defined feature number
ARCIDNUM	Stream Segment Number (Unique Identifier used to link all tables and coverages)
HECTARES2	Area of individual stream segmentshed in hectares
SHED_INCLUS	Area of watershed in hectares
STRAHLER	Strahler Stream Order
LINK	Shreve Link
GEO2W_DEV	Percent of watershed classified System Geology Type Devonian
GEO2W_MSPI	Percent of watershed classified System Geology Type Mississippian
GEO2W_ORD	Percent of watershed classified System Geology Type Ordovician
GEO2W_PENN	Percent of watershed classified System Geology Type Pennsylvanian
GEO2W_QUAT	Percent of watershed classified System Geology Type Quaternary
GEO2W_SILR	Percent of watershed classified System Geology Type Silurian
HGW_B	Percent of watershed classified as Hydrologic Soil Group B in STATSGO dataset
HGW_C	Percent of watershed classified as Hydrologic Soil Group C in STATSGO dataset
HGW_D	Percent of watershed classified as Hydrologic Soil Group D in STATSGO dataset
SF2W_CHRT	Percent of watershed classified as Cherty (combination of STATSGO CRSIL, CRVSL)
SF2W_CLAY	Percent of watershed classified as Clays (combination of STATSGO CL, SIC, SICL)
SF2W_LOAM	Percent of watershed classified as Loams (combination of STATSGO L, SIL, VAR)
SF2W_SAND	Percent of watershed classified as Sandy (combination of STATSGO FSL, LS)
SF2W_STONY	Percent of watershed classified as Stony (combination of STATSGO STL, STSIL)
RELCATW_1	Percent of watershed in Relief Category 1 (0-50 feet)
RELCATW_2	Percent of watershed in Relief Category 2 (51-100 feet)
RELCATW_3	Percent of watershed in Relief Category 3 (101-200 feet)
RELCATW_4	Percent of watershed in Relief Category 4 (201-300 feet)
RELCATW_5	Percent of watershed in Relief Category 5 (301-500 feet)
RELCATW_6	Percent of watershed in Relief Category 6 (501-700 feet)
SLOP_WAT	Mean slope of watershed
OUT	Identifies stream segements that fall outside of Missouri and will not have landscape attributes
OUT_ABOVE	Summarizes the number of streams above each segment that are outside of Missouri
HW_POND	Indicates presence/absence of an impoundment (size 2-15 acres) on a stream segment
HWP_ABOVE	Number of impoundments (size 2-15 acres) above segment
HWPT	Indicates presence/absence of impoundment (size 2-15 acres) within a 100 meter tolerance
HWPT_ABOVE	Number of impoudments (size 2-15 acres) within a 100 meter tolerance above each segment
HWP2	Indicates presence/absence of an impoundment (size 2-50 acres) on a stream segment
HWP2_ABOVE	Number of impoundments (size 2-50 acres) above segment
HWP2T	Indicates presence/absence of impoundment (size 2-50 acres) within a 100 meter tolerance
HWP2T_ABOVE	Number of impoudments (size 2-50 acres) within a 100 meter tolerance above each segment
WB_INT	Indicates presence/absence of any size waterbody intersecting the stream segment

Appendix D. Attribute descriptions for the Reduced Set of watershed landscape variables used in the cluster analyses. Corresponds with the CLUSTRED_VARS coverage.

Attribute	Description
FNODE#	Internal node number for the beginning of an arc (from-node)
TNODE#	Internal node number for the end of an arc (to-node)
LPOLY#	Internal node number for the left polygon
RPOLY#	Internal node number for the right polygon
LENGTH	Length (in meters) of stream segment
CLUSTRED_VARS#	Internal feature number. Sequential unique whole numbers that are automatically generated.
CLUSTRED_VARS-ID	User defined feature number
ARCIDNUM	Stream Segment Number (Unique Identifier used to link all tables and coverages)
HECTARES2	Area of individual stream segmentshed in hectares
SHED_INCLUS	Area of watershed in hectares
STRAHLER	Strahler Stream Order
LINK	Shreve Link
HGW_B	Percent of watershed classified as Hydrologic Soil Group B in STATSGO dataset
HGW_C	Percent of watershed classified as Hydrologic Soil Group C in STATSGO dataset
HGW_D	Percent of watershed classified as Hydrologic Soil Group D in STATSGO dataset
RELCATW_1	Percent of watershed in Relief Category 1 (0-50 feet)
RELCATW_2	Percent of watershed in Relief Category 2 (51-100 feet)
RELCATW_3	Percent of watershed in Relief Category 3 (101-200 feet)
RELCATW_4	Percent of watershed in Relief Category 4 (201-300 feet)
RELCATW_5	Percent of watershed in Relief Category 5 (301-500 feet)
RELCATW_6	Percent of watershed in Relief Category 6 (501-700 feet)
SLOP_WAT	Mean slope of watershed
OUT	Identifies stream segements that fall outside of Missouri and will not have landscape attributes
OUT_ABOVE	Summarizes the number of streams above each segment that are outside of Missouri
HW_POND	Indicates presence/absence of an impoundment (size 2-15 acres) on a stream segment
HWP_ABOVE	Number of impoundments (size 2-15 acres) above segment
HWPT	Indicates presence/absence of impoundment (size 2-15 acres) within a 100 meter tolerance
HWPT_ABOVE	Number of impoudments (size 2-15 acres) within a 100 meter tolerance above each segment
HWP2	Indicates presence/absence of an impoundment (size 2-50 acres) on a stream segment
HWP2_ABOVE	Number of impoundments (size 2-50 acres) above segment
HWP2T	Indicates presence/absence of impoundment (size 2-50 acres) within a 100 meter tolerance
HWP2T_ABOVE	Number of impoudments (size 2-50 acres) within a 100 meter tolerance above each segment
WB_INT	Indicates presence/absence of any size waterbody intersecting the stream segment

Appendix E. Attribute descriptions for the Watershed Land Cover variables used in the cluster analyses. Corresponds with the CLUST_LULCW_VARS coverage.

Attribute	Description
FNODE#	Internal node number for the beginning of an arc (from-node)
TNODE#	Internal node number for the end of an arc (to-node)
LPOLY#	Internal node number for the left polygon
RPOLY#	Internal node number for the right polygon
LENGTH	Length (in meters) of stream segment
CL_LULCW_VARS#	Internal feature number. Sequential unique whole numbers that are automatically generated.
CL_LULCW_VARS-ID	User defined feature number
ARCIDNUM	Stream Segment Number (Unique Identifier used to link all tables and coverages)
HECTARES2	Area of individual stream segmentshed in hectares
SHED_INCLUS	Area of watershed in hectares
STRAHLER	Strahler Stream Order
LINK	Shreve Link
LULCW_CROP	Percent of watershed classified as Cropland
LULCW_FOR	Percent of watershed classified as Forest
LULCW_GRASS	Percent of watershed classified as Grassland
LULCW_SWAMP	Percent of watershed classified as Swamp
LULCW_URB	Percent of watershed classified as Urban
LULCW_WAT	Percent of watershed classified as Water
OUT	Identifies stream segements that fall outside of Missouri and will not have landscape attributes
OUT_ABOVE	Summarizes the number of streams above each segment that are outside of Missouri
HW_POND	Indicates presence/absence of an impoundment (size 2-15 acres) on a stream segment
HWP_ABOVE	Number of impoundments (size 2-15 acres) above segment
HWPT	Indicates presence/absence of impoundment (size 2-15 acres) within a 100 meter tolerance
HWPT_ABOVE	Number of impoudments (size 2-15 acres) within a 100 meter tolerance above each segment
HWP2	Indicates presence/absence of an impoundment (size 2-50 acres) on a stream segment
HWP2_ABOVE	Number of impoundments (size 2-50 acres) above segment
HWP2T	Indicates presence/absence of impoundment (size 2-50 acres) within a 100 meter tolerance
HWP2T_ABOVE	Number of impoudments (size 2-50 acres) within a 100 meter tolerance above each segment
WB_INT	Indicates presence/absence of any size waterbody intersecting the stream segment

Appendix F. Attribute descriptions for the Local (segmentshed) Land Cover variables used in the cluster analyses. Corresponds with the CLUST_LULCL_VARS coverage.

Attribute	Description
FNODE#	Internal node number for the beginning of an arc (from-node)
TNODE#	Internal node number for the end of an arc (to-node)
LPOLY#	Internal node number for the left polygon
RPOLY#	Internal node number for the right polygon
LENGTH	Length (in meters) of stream segment
CL_LULCL_VARS#	Internal feature number. Sequential unique whole numbers that are automatically generated.
CL_LULCL_VARS-ID	User defined feature number
ARCIDNUM	Stream Segment Number (Unique Identifier used to link all tables and coverages)
HECTARES2	Area of individual stream segmentshed in hectares
SHED_INCLUS	Area of watershed in hectares
STRAHLER	Strahler Stream Order
LINK	Shreve Link
LULCL_CROP	Percent of local segmentshed classified as Cropland
LULCL_FOR	Percent of local segmentshed classified as Forest
LULCL_GRASS	Percent of local segmentshed classified as Grassland
LULCL_SWAMP	Percent of local segmentshed classified as Swamp
LULCL_URB	Percent of local segmentshed classified as Urban
LULCL_WAT	Percent of local segmentshed classified as Water
OUT	Identifies stream segements that fall outside of Missouri and will not have landscape attributes
OUT_ABOVE	Summarizes the number of streams above each segment that are outside of Missouri
HW_POND	Indicates presence/absence of an impoundment (size 2-15 acres) on a stream segment
HWP_ABOVE	Number of impoundments (size 2-15 acres) above segment
HWPT	Indicates presence/absence of impoundment (size 2-15 acres) within a 100 meter tolerance
HWPT_ABOVE	Number of impoudments (size 2-15 acres) within a 100 meter tolerance above each segment
HWP2	Indicates presence/absence of an impoundment (size 2-50 acres) on a stream segment
HWP2_ABOVE	Number of impoundments (size 2-50 acres) above segment
HWP2T	Indicates presence/absence of impoundment (size 2-50 acres) within a 100 meter tolerance
HWP2T_ABOVE	Number of impoudments (size 2-50 acres) within a 100 meter tolerance above each segment
WB_INT	Indicates presence/absence of any size waterbody intersecting the stream segment

Appendix G. Attribute descriptions for cluster results (numbers and distance ranks) generated for the Full Set of watershed variables. Corresponds with the CLUST_ALL coverage.

Attribute	Description
FNODE#	Internal node number for the beginning of an arc (from-node)
TNODE#	Internal node number for the end of an arc (to-node)
LPOLY#	Internal node number for the left polygon
RPOLY#	Internal node number for the right polygon
LENGTH	Length (in meters) of stream segment
CLUST_ALL#	Internal feature number. Sequential unique whole numbers that are automatically generated.
CLUST_ALL-ID	User defined feature number
ARCIDNUM	Stream Segment Number (Unique Identifier used to link all tables and coverages)
HECTARES2	Area of individual stream segmentshed in hectares
SHED_INCLUS	Area of watershed in hectares
STRAHLER	Strahler Stream Order
LINK	Shreve Link
ALL_CL2	Full set 2 clusters
ALL_D2	Distance ranks for 2 clusters
ALL_CL4	Full set 4 clusters
ALL_D4	Distance ranks for 4 clusters
ALL_CL6	Full set 6 clusters
ALL_D6	Distance ranks for 6 clusters
ALL_CL8	Full set 8 clusters
ALL_D8	Distance ranks for 8 clusters
ALL_CL10	Full set 10 clusters
ALL_D10	Distance ranks for 10 clusters
ALL_CL12	Full set 12 clusters
ALL_D12	Distance ranks for 12 clusters
ALL_CL14	Full set 14 clusters
ALL_D14	Distance ranks for 14 clusters
ALL_CL16	Full set 16 clusters
ALL_D16	Distance ranks for 16 clusters
ALL_CL18	Full set 18 clusters
ALL_D18	Distance ranks for 18 clusters
ALL_CL20	Full set 20 clusters
ALL_D20	Distance ranks for 20 clusters
OUT	Identifies stream segements that fall outside of Missouri and will not have landscape attributes
OUT_ABOVE	Summarizes the number of streams above each segment that are outside of Missouri
HW_POND	Indicates presence/absence of an impoundment (size 2-15 acres) on a stream segment
HWP_ABOVE	Number of impoundments (size 2-15 acres) above segment
HWPT	Indicates presence/absence of impoundment (size 2-15 acres) within a 100 meter tolerance
HWPT_ABOVE	Number of impoudments (size 2-15 acres) within a 100 meter tolerance above each segment
HWP2	Indicates presence/absence of an impoundment (size 2-50 acres) on a stream segment
HWP2_ABOVE	Number of impoundments (size 2-50 acres) above segment
HWP2T	Indicates presence/absence of impoundment (size 2-50 acres) within a 100 meter tolerance
HWP2T_ABOVE	Number of impoudments (size 2-50 acres) within a 100 meter tolerance above each segment
WB_INT	Indicates presence/absence of any size waterbody intersecting the stream segment

Appendix H. Attribute descriptions for cluster results (numbers and distance ranks) generated for the Reduced Set of watershed variables. Corresponds with the CLUST_RED coverage.

Attribute	Description
FNODE#	Internal node number for the beginning of an arc (from-node)
TNODE#	Internal node number for the end of an arc (to-node)
LPOLY#	Internal node number for the left polygon
RPOLY#	Internal node number for the right polygon
LENGTH	Length (in meters) of stream segment
CLUST_RED#	Internal feature number. Sequential unique whole numbers that are automatically generated.
CLUST_RED-ID	User defined feature number
ARCIDNUM	Stream Segment Number (Unique Identifier used to link all tables and coverages)
HECTARES2	Area of individual stream segmentshed in hectares
SHED_INCLUS	Area of watershed in hectares
STRAHLER	Strahler Stream Order
LINK	Shreve Link
RED_CL2	Reduced set 2 clusters
RED_D2	Distance ranks for 2 cluster set
RED_CL4	Reduced set 4 clusters
RED_D4	Distance ranks for 4 cluster set
RED_CL6	Reduced set 6 clusters
RED_D6	Distance ranks for 6 cluster set
RED_CL8	Reduced set 8 clusters
RED_D8	Distance ranks for 8 cluster set
RED_CL10	Reduced set 10 clusters
RED_D10	Distance ranks for 10 cluster set
RED_CL12	Reduced set 12 clusters
RED_D12	Distance ranks for 12 cluster set
RED_CL14	Reduced set 14 clusters
RED_D14	Distance ranks for 14 cluster set
RED_CL16	Reduced set 16 clusters
RED_D16	Distance ranks for 16 cluster set
RED_CL18	Reduced set 18 clusters
RED_D18	Distance ranks for 18 cluster set
RED_CL20	Reduced set 20 clusters
RED_D20	Distance ranks for 20 cluster set
OUT	Identifies stream segements that fall outside of Missouri and will not have landscape attributes
OUT_ABOVE	Summarizes the number of streams above each segment that are outside of Missouri
HW_POND	Indicates presence/absence of an impoundment (size 2-15 acres) on a stream segment
HWP_ABOVE	Number of impoundments (size 2-15 acres) above segment
HWPT	Indicates presence/absence of impoundment (size 2-15 acres) within a 100 meter tolerance
HWPT_ABOVE	Number of impoudments (size 2-15 acres) within a 100 meter tolerance above each segment
HWP2	Indicates presence/absence of an impoundment (size 2-50 acres) on a stream segment
HWP2_ABOVE	Number of impoundments (size 2-50 acres) above segment
HWP2T	Indicates presence/absence of impoundment (size 2-50 acres) within a 100 meter tolerance
HWP2T_ABOVE	Number of impoudments (size 2-50 acres) within a 100 meter tolerance above each segment
WB_INT	Indicates presence/absence of any size waterbody intersecting the stream segment

Appendix I. Attribute descriptions for cluster results (numbers and distance ranks) generated for the Watershed Land Cover variables. Corresponds with the CLUST_LULCW coverage.

Attribute	Description
FNODE#	Internal node number for the beginning of an arc (from-node)
TNODE#	Internal node number for the end of an arc (to-node)
LPOLY#	Internal node number for the left polygon
RPOLY#	Internal node number for the right polygon
LENGTH	Length (in meters) of stream segment
CLUST_LULCW#	Internal feature number. Sequential unique whole numbers that are automatically generated.
CLUST_LULCW-ID	User defined feature number
ARCIDNUM	Stream Segment Number (Unique Identifier used to link all tables and coverages)
HECTARES2	Area of individual stream segmentshed in hectares
SHED_INCLUS	Area of watershed in hectares
STRAHLER	Strahler Stream Order
LINK	Shreve Link
LULCW_CL2	Watershed Land Cover 2 Clusters
LULCW_D2	Distance ranks for watershed land cover 2 clusters
LULCW_CL4	Watershed Land Cover 4 Clusters
LULCW_D4	Distance ranks for watershed land cover 4 clusters
LULCW_CL6	Watershed Land Cover 6 Clusters
LULCW_D6	Distance ranks for watershed land cover 6 clusters
LULCW_CL8	Watershed Land Cover 8 Clusters
LULCW_D8	Distance ranks for watershed land cover 8 clusters
LULCW_CL10	Watershed Land Cover 10 Clusters
LULCW_D10	Distance ranks for watershed land cover 10 clusters
LULCW_CL12	Watershed Land Cover 12 Clusters
LULCW_D12	Distance ranks for watershed land cover 12 clusters
LULCW_CL14	Watershed Land Cover 14 Clusters
LULCW_D14	Distance ranks for watershed land cover 14 clusters
LULCW_CL16	Watershed Land Cover 16 Clusters
LULCW_D16	Distance ranks for watershed land cover 16 clusters
LULCW_CL18	Watershed Land Cover 18 Clusters
LULCW_D18	Distance ranks for watershed land cover 18 clusters
LULCW_CL20	Watershed Land Cover 20 Clusters
LULCW_D20	Distance ranks for watershed land cover 20 clusters
OUT	Identifies stream segements that fall outside of Missouri and will not have landscape attributes
OUT_ABOVE	Summarizes the number of streams above each segment that are outside of Missouri
HW_POND	Indicates presence/absence of an impoundment (size 2-15 acres) on a stream segment
HWP_ABOVE	Number of impoundments (size 2-15 acres) above segment
HWPT	Indicates presence/absence of impoundment (size 2-15 acres) within a 100 meter tolerance
HWPT_ABOVE	Number of impoudments (size 2-15 acres) within a 100 meter tolerance above each segment
HWP2	Indicates presence/absence of an impoundment (size 2-50 acres) on a stream segment
HWP2_ABOVE	Number of impoundments (size 2-50 acres) above segment
HWP2T	Indicates presence/absence of impoundment (size 2-50 acres) within a 100 meter tolerance
HWP2T_ABOVE	Number of impoudments (size 2-50 acres) within a 100 meter tolerance above each segment
WB_INT	Indicates presence/absence of any size waterbody intersecting the stream segment

Appendix J. Attribute descriptions for cluster results (numbers and distance ranks) generated for the Local Land Cover variables. Corresponds with the CLUST_LULCL coverage.

Attribute	Description
FNODE#	Internal node number for the beginning of an arc (from-node)
TNODE#	Internal node number for the end of an arc (to-node)
LPOLY#	Internal node number for the left polygon
RPOLY#	Internal node number for the right polygon
LENGTH	Length (in meters) of stream segment
CLUST_LULCW#	Internal feature number. Sequential unique whole numbers that are automatically generated.
CLUST_LULCW-ID	User defined feature number
ARCIDNUM	Stream Segment Number (Unique Identifier used to link all tables and coverages)
HECTARES2	Area of individual stream segmentshed in hectares
SHED_INCLUS	Area of watershed in hectares
STRAHLER	Strahler Stream Order
LINK	Shreve Link
LULCW_CL2	Local segmentshed Land Cover 2 Clusters
LULCW_D2	Distance ranks for local segmentshed land cover 2 clusters
LULCW_CL4	Local segmentshed Land Cover 4 Clusters
LULCW_D4	Distance ranks for local segmentshed land cover 4 clusters
LULCW_CL6	Local segmentshed Land Cover 6 Clusters
LULCW_D6	Distance ranks for local segmentshed land cover 6 clusters
LULCW_CL8	Local segmentshed Land Cover 8 Clusters
LULCW_D8	Distance ranks for local segmentshed land cover 8 clusters
LULCW_CL10	Local segmentshed Land Cover 10 Clusters
LULCW_D10	Distance ranks for local segmentshed land cover 10 clusters
LULCW_CL12	Local segmentshed Land Cover 12 Clusters
LULCW_D12	Distance ranks for local segmentshed land cover 12 clusters
LULCW_CL14	Local segmentshed Land Cover 14 Clusters
LULCW_D14	Distance ranks for local segmentshed land cover 14 clusters
LULCW_CL16	Local segmentshed Land Cover 16 Clusters
LULCW_D16	Distance ranks for local segmentshed land cover 16 clusters
LULCW_CL18	Local segmentshed Land Cover 18 Clusters
LULCW_D18	Distance ranks for local segmentshed land cover 18 clusters
LULCW_CL20	Local segmentshed Land Cover 20 Clusters
LULCW_D20	Distance ranks for local segmentshed land cover 20 clusters
OUT	Identifies stream segements that fall outside of Missouri and will not have landscape attributes
OUT_ABOVE	Summarizes the number of streams above each segment that are outside of Missouri
HW_POND	Indicates presence/absence of an impoundment (size 2-15 acres) on a stream segment
HWP_ABOVE	Number of impoundments (size 2-15 acres) above segment
HWPT	Indicates presence/absence of impoundment (size 2-15 acres) within a 100 meter tolerance
HWPT_ABOVE	Number of impoudments (size 2-15 acres) within a 100 meter tolerance above each segment
HWP2	Indicates presence/absence of an impoundment (size 2-50 acres) on a stream segment
HWP2_ABOVE	Number of impoundments (size 2-50 acres) above segment
HWP2T	Indicates presence/absence of impoundment (size 2-50 acres) within a 100 meter tolerance
HWP2T_ABOVE	Number of impoudments (size 2-50 acres) within a 100 meter tolerance above each segment
WB_INT	Indicates presence/absence of any size waterbody intersecting the stream segment